PHYSICAL SCIENCE STUDY COMMITTEE

COLLEGE PHYSICS

COLLEGE PHYSICS

COLLEGE PHYSICS

PHYSICAL SCIENCE STUDY COMMITTEE

RAYTHEON EDUCATION COMPANY

The materials taken from the original and second editions and the Advanced Topics of PSSC PHYSICS included in this text will be available to all publishers for use in English after December 31, 1970, and in translations after December 31, 1975. Publishers desiring to use these materials are referred to the text PSSC PHYSICS.

RAYTHEON

LIBRARY OF CONGRESS CATALOG CARD NUMBER 68-17716

Published by RAYTHEON EDUCATION COMPANY

Under arrangements with Education Development Center, Inc.

Copyright © 1968 Education Development Center, Inc.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage or retrieval system, without permission in writing from the publisher.

Printed in the United States of America

Printed January, 1968

PREFACE

It has been remarked by many college professors over the last few years that the PSSC course in physics could serve very well the needs of many college freshmen, provided certain changes in content and emphasis were made. These changes must take into account the fact that curriculum reform in science in the junior and senior high school is now in high gear, and that, therefore, even the student who did not take physics in high school and does not intend to major in science or engineering comes to college with a better grasp of scientific inquiry than was previously the case. Furthermore, the college freshman who takes physics (either because he plans to major in science or engineering, or because he elects it for its cultural value) now expects to get from his work a sound understanding of the principles of physics, including the physics of the twentieth century. Much of the content of PSSC College Physics and its accompanying Laboratory Guide comes from the second edition of the PSSC course. Substantial parts were added and material from the earlier edition was deleted. Some of the new material was published before as the Advanced Topics Supplement to the PSSC course. Other parts were written specifically for this book. Besides additions and deletions, several chapters were rearranged for greater coherence. The end-of-chapter material contains many easy problems, as well as more challenging questions designed for the sharper student.

Although the text contains many photographs and detailed descriptions of experiments, there is no substitute for the students' direct investigations in the laboratory. Taken as a whole, the text and Laboratory Guide are designed in content and emphasis to serve as the backbone for a mature, deep, and modern course in physics both for those who have studied some physics before and those who have not.

ACKNOWLEDGEMENTS

The writing of new material for this edition, as well as the editorial changes, was done by Judson B. Cross, John H. Dodge, James A. Walter and myself of the EDC staff. We were aided by the work of Ingemar Bartholdson, Stockholm; K. G. Friskopp, Örebro, and Goren Lindahl, Västerås, both in Sweden and here. The production was supervised by Benjamin T. Richards. The responsibility for the selection of the material and its organization is mine.

In a very direct way we built on the work of those responsible for the second edition of PSSC PHYSICS and the Advanced Topics Supplement. In addition to the EDC staff these people were: Professors Robert Estin, Roosevelt University; Bruno Ferretti, University of Bologna, Italy; Kerson Huang, M.I.T.; Arthur Kerman, M.I.T.; Derek Livesey, University of British Columbia; Stefan Machlup, Western Reserve University, Philip Morrison, M.I.T.; Melba Phillips, University of Chicago; and Byron L. Youtz, Reed College, editor of the second edition.

It is obvious that this edition could not have become a reality without the effort of the original members of the Physical Science Study Committee, under the leadership of Prof. Jerrold R. Zacharias and the late Prof. Francis L. Friedman.

Uri Haber-Schaim

Newton, Massachusetts
September 1967

CONTENTS

chapter 1 **Time, Space, and Matter** 1

1-1 Time and Its Measurement 1-2 Space and Its Measurement 1-3 Matter 1-4 On the Limitations of Measuring 1-5 Fractional Error 1-6 Black Boxes

chapter 2 **Data, Graphs, and Functions** 19

2-1 Tables and Graphs 2-2 Direct Proportion and Linear Functions 2-3 Power Laws and Similar Figures 2-4 The Inverse-Square Relation and Stellar Distances 2-5 The Exponential Function 2-6 The Exponential Function and Semi-Log Plots 2-7 Power Laws and Log-Log Plots 2-8 Scaling – The Physics of Lilliput

chapter 3 **The Propagation of Light** 43

3-1 Sources of Light 3-2 Transparent, Colored, and Opaque Materials 3-3 Reflection 3-4 Light-Sensitive Devices 3-5 Invisible Light 3-6 How Light Travels 3-7 The Speed of Light 3-8 Light Beams, Pencils, and Rays 3-9 How We Locate Objects

chapter 4 **Reflection and Refraction** 54

4-1 The Laws of Reflection 4-2 Images in Plane Mirrors 4-3 Refraction 4-4 Experiments on the Angles in Refraction 4-5 The Index of Refraction: Snell's Law 4-6 The Absolute Index of Refraction 4-7 The Passage of Light from Glass (or water) to Air: Reversibility 4-8 The Passage of Light from Glass to Water 4-9 Total Internal Reflection 4-10 Refraction by Prisms; Dispersion

chapter 5 **The Particle Model of Light** 75

5-1 Reflection 5-2 Refraction 5-3 Source Strength and Intensity of Illumination 5-4 Light Pressure 5-5 Absorption and Heating 5-6 Some Difficulties with the Particle Theory 5-7 The Speed of Light and the Theory of Refraction 5-8 The Status of the Particle Model

chapter 6 **Introduction to Waves** 87

6-1 A Wave: Something Else that Travels 6-2 Waves on Coil Springs 6-3 Superposition: Pulses Crossing 6-4 Reflection and Transmission 6-5 Idealizations and Approximations 6-6 A Wave Model for Light?

chapter 7 **Waves and Light** 100

7-1 Water Waves 7-2 Straight and Circular Pulses 7-3 Reflection 7-4 Speed of Propagation and Periodic Waves 7-5 Refraction 7-6 Dispersion 7-7 Diffraction

chapter 8 **Interference** 114

8-1 Interference on a Spring 8-2 Interference from Two Point Sources 8-3 The Shape of Nodal Lines 8-4 Wavelengths, Source Separation, and Angles 8-5 Phase 8-6 Summary and Conclusion

chapter 9 **Light Waves** 127

9-1 Can We See Interference in Light? 9-2 Interference of Light Waves: Young's Experiment 9-3 The Phase of Light Sources: Atoms 9-4 Color and Wavelength of Light 9-5 Diffraction: An Interference Effect in Single Slits? 9-6 A Theory of Diffraction by a Slit 9-7 Experimental Checks with Single and Double Slits 9-8 Interference in Thin Films 9-9 Interference in Light Transmitted Through Thin Films 9-10 Color Effects in Interference 9-11 Conclusion

chapter 10 **Applied Optics** 144

10-1 Parabolic Mirrors 10-2 Searchlights 10-3 Astronomical Telescopes 10-4 The Use of Astronomical Telescopes 10-5 The Convergence of Light by a Set of Prisms 10-6 Lenses 10-7 Images Formed by Lenses 10-8 Real and Virtual Images 10-9 The Magnifier or Simple Microscope 10-10 The Compound Microscope 10-11 Refracting Telescopes 10-12 Limitations of Optical Instruments: Aberrations and Resolutions

chapter 11 **Motion Along a Straight-line Path** 169

11-1 Position and Displacement along a Line 11-2 Velocity 11-3 Varying Velocity 11-4 Analysis of Position-Time Graphs 11-5 Slope 11-6 Instantaneous Velocity—The Slope of the Tangent Line 11-7 Acceleration 11-8 Useful Results for Motion with Constant Acceleration 11-9 Summary

chapter 12 **Motion in Space** 195

12-1 Position and Displacement 12-2 Addition and Subtraction of Displacements 12-3 Multiplying Vectors by Numbers and Scalars 12-4 Velocity Vectors in Space 12-5 Components of Vectors 12-6 Velocity Changes and Constant Vector Acceleration 12-7 Changing Acceleration 12-8 The Description of Motion; Frames of Reference 12-9 Kinematics and Dynamics 12-10 The Speed of Light

chapter 13 **Newton's Law of Motion** 218

13-1 Ideas about Force and Motion 13-2 Motion Without Force 13-3 Changes in Velocity when a Constant Force Acts 13-4 Dependence of Change of Velocity on Magnitude of Force 13-5 Inertial Mass 13-6 Inertial and Gravitational Mass 13-7 Newton's Law: Dynamical Measurement of Force; Units 13-8 Forces that Change, and Newton's Law

13-9 How Forces Add; The Net Force 13-10 The Vector Nature of Newton's law 13-11 Forces in Nature

chapter 14 **Motion at the Earth's Surface** 235

14-1 Weight and the Gravitational Field of the Earth 14-2 Free Fall 14-3 Projectile Motion: The Vector Nature of Newton's Law of Motion 14-4 Projectile Motion: Determination of the Path 14-5 Deflecting Forces and Circular Motion 14-6 Earth Satellites 14-7 The Moon's Motion 14-8 Simple Harmonic Motion 14-9 Experimental Frames of Reference 14-10 Fictitious Forces in Accelerated Frames 14-11 Newton's Law and the Rotation of the Earth

chapter 15 **Universal Gravitation and the Solar System** 258

15-1 Early Planetary Systems 15-2 Copernicus' Planetary System 15-3 Objections to Copernicus' Theory 15-4 Tycho Brahe 15-5 Kepler 15-6 Kinematic Description and the Dynamical Problem 15-7 Newton 15-8 Universal Gravitation 15-9 Some of Newton's Later Accomplishments 15-10 Laboratory Tests of the Law of Universal Gravitation 15-11 A Small Discrepancy

chapter 16 **Momentum and the Conservation of Momentum** 278

16-1 Impulse 16-2 Momentum 16-3 Changes in Momentum when Two Bodies Interact 16-4 The Law of Conservation of Momentum 16-5 The Center of Mass 16-6 The Conservation of Momentum in General 16-7 Forces of Interaction

chapter 17 **Work and Kinetic Energy** 300

17-1 Energy Transfer 17-2 Work: A Measure of Energy Transfer 17-3 More About the Definition of Work 17-4 Kinetic Energy 17-5 The Transfer of Kinetic Energy from One Mass to Another 17-6 Another Look at the Simple Collision 17-7 Conservation of Kinetic Energy in Elastic Interactions 17-8 Kinetic Energy and Momentum 17-9 Work and Kinetic Energy When More Than One Force Acts 17-10 Loss of Kinetic Energy in a Frictional Interaction 17-11 Conclusion

chapter 18 **Potential Energy** 320

18-1 The Spring Bumper 18-2 Potential Energy of Two Interacting Bodies 18-3 Gravitational Potential Energy near the Surface of the Earth 18-4 Gravitional Potential Energy in General 18-5 Escape Energy and Binding Energy 18-6 Total Mechanical Energy

chapter 19 **Angular Momentum and Its Conservation** 341

19-1 The Law of Equal Areas 19-2 Angular Momentum 19-3 Energy, Angular Momentum, and Trajectories 19-4 A Special Case: The Motion of Satellites 19-5 Angular Momentum: A Vector? 19-6 Conservation of Angular Momentum for Interacting Bodies 19-7 The Rotation of Rigid Bodies 19-8 The Angular Momentum of Rigid Bodies; Moment of Inertia 19-9 Torque—The Rate of Change of Angular Momentum 19-10 Orbital Angular Momentum and Spin 19-11 Conservation of Angular Momentum in General

chapter 20 **Heat, Molecular Motion, and Conservation of Energy** 368

20-1 Gas Pressure, Boltzmann's Constant, and Molecules 20-2 The Dynamics of Gases 20-3 Temperature and Molecular Kinetic Theory; Internal Energy 20-4 Mechanical Energy of Bulk Motion and Internal Energy 20-5 A Quantitative Study of the Conversion of Mechanical to Internal Energy 20-6 Heat Flow 20-7 Quantitative Relation of Energy Dissipation and Temperature Rise 20-8 Conservation of Energy

chapter 21 **Irreversible Processes** 384

21-1 Some Examples 21-2 An Experiment with Marbles 21-3 Qualitative Explanation of the Marble Experiment 21-4 Some Basic Ideas about Probability 21-5 States and Distributions 21-6 The Most Likely Distribution 21-7 Free Expansion of a Gas 21-8 Quantitative Aspects 21-9 Density Fluctuations 21-10 Inelastic Collisions and Heat Conduction

chapter 22 **Entropy** 402

22-1 The Macroscopic State 22-2 Reversible Process 22-3 The Gas Oscillator 22-4 Adiabatic and Isothermal Processes 22-5 Free Expansion and Isothermal Expansion of an Ideal Gas 22-6 Entropy 22-7 A Realistic Heat Bath 22-8 Entropy Changes in an Inelastic Collision 22-9 Entropy Changes in Heat Conduction at Constant Volume 22-10 The Entropy of an Ideal Gas When Both Volume and Temperature Change 22-11 Diffusion – The Entropy of a Perfect Gas 22-12 The Second Law of Thermodynamics

chapter 23 **Some Qualitative Facts About Electricity** 420

23-1 Attraction and Repulsion between Electrified Objects 23-2 Electric Forces between the Building Blocks of Matter 23-3 Insulators and Conductors 23-4 Some Experiments with an Electroscope 23-5 Electrostatic Induction 23-6 Improved Electroscopes and Electrometers 23-7 Batteries 23-8 Electric Currents 23-9 Conductivity of Gases; Ionization 23-10 The Cloud Chamber 23-11 Conductivity of Solutions 23-12 Electrons in Metals 23-13 Diodes, Electron Guns, Cathode-Ray Oscilloscopes

chapter 24 **Coulomb's Law and the Elementary Electric Charge** 438

24-1 Force vs. Distance 24-2 Electric Charge and Electric Force 24-3 Electric Force Fields 24-4 Measuring Small Electric Forces 24-5 The Elementary Charge 24-6 A Large Electric Balance 24-7 The Constant in Coulomb's Law 24-8 Conservation of Charge 24-9 The Electric Charge of Electrons and Other Particles of Matter

chapter 25 **Energy and Motion of Charges in Electric Fields** 461

25-1 Determining the Mass of the Electron and of the Proton 25-2 Electric Current 25-3 Electrolytic Measurement of Electric Currents 25-4 Experimental Checks on Energy Transfer, Electric Forces, and Elementary Charges 25-5 EMF and the Energy Supplied by a Battery 25-6

Electric Field and Electric Potential 25-7 Batteries, Volts, and Amperes 25-8 Status Report

chapter 26 **The Rutherford Atom** 483

26-1 The Deflection of Alpha Particles and the Rutherford Model of Atoms 26-2 The Trajectories of Alpha Particles in the Electric Field of a Nucleus 16-3 Angular Distribution of Scattering 26-4 More Information from Scattering

chapter 27 **Electric Circuits** 497

27-1 Conductors, Batteries, and Potential Difference 27-2 Measurement of Potential Difference 27-3 A Further Check on Potential Difference and Energy 27-4 Current Versus Potential Difference 27-5 An Overall View of a Circuit

chapter 28 **The Magnetic Field** 514

28-1 The Magnetic Needle 28-2 Magnetic Fields of Magnets and Currents 28-3 The Vector Addition of Magnetic Fields 28-4 Forces on Currents in Magnetic Fields—A Unit of Magnetic Field Strength 28-5 Meters and Motors 28-6 Magnetic Field near a Long Straight Wire 28-7 Circulation 28-8 Uniform Magnetic Fields 28-9 Forces on Moving Charged Particles in a Magnetic Field 28-10 Using Magnetic Fields to Measure the Masses of Charged Particles 28-11 What Alpha Particles Are

chapter 29 **Electromagnetic Induction and Electromagnetic Waves** 540

29-1 Induced Current 29-2 Relative Motion 29-3 Magnetic Flux Change 29-4 Induced EMF 29-5 Direction of the Induced EMF 29-6 Electric Fields around Changing Magnetic Fluxes 29-7 Magnetic Fields around Changing Electric Fluxes 29-8 The Mechanism of Electromagnetic Radiation 29-9 Evidence for Electromagnetic Radiation; The Electromagnetic Spectrum

chapter 30 **Relativistic Kinematics** 562

30-1 A Medium for Light Waves? 30-2 Waves in Moving Media 30-3 A One-way Measurement of the Speed of Light 30-4 The Michelson-Morley Experiment 30-5 The Lorentz Contraction 30-6 Twin-source Interference in a Moving Medium 30-7 A First-order Experiment with Light 30-8 The Speed of Light: A Universal Constant 30-9 The Fizeau Experiment 39-10 The Relativistic Addition of Velocities 30-11 An Experimental Check 30-12 Time for Different Observers 30-13 The Lorentz Transformation 30-14 Time Dilation 30-15 An Experiment with Muons 30-16 The Length of Moving Objects 30-17 Summary Appendix: The Twin Paradox

chapter 31 **Photons** 589

31-1 The Graininess of Light 31-2 The Photoelectric Effect 31-3 Einstein's Interpretation of the Photoelectric Effect 31-4 The Orderliness of Chance 31-5 Graininess and Interference 31-6 Photons and Electromagnetic Waves chapter 32 **Speed, Energy, and Mass** 604

32-1 The Ultimate Speed 32-2 Speed and Kinetic Energy 32-3 Momentum 32-4 The Momentum of Photons: The Pressure of Light 32-5 Compton Scattering 32-6 The Electron-Positron Annihilation 32-7 Nuclear Reactions: Conservation of Total Energy 32-8 Mass and Energy 32-9 Mass for Photons? 32-10 Review and Outlook

chapter 33 **Atoms and Spectra** 624

33-1 The Stability of Atoms 33-2 The Experiments of Franck and Hertz; Atomic Energy Levels 33-3 Dissecting Atomic Spectra: Excitation and Emission 33-4 Absorption Spectra 33-5 The Energy Levels of Hydrogen

chapter 34 **Matter Waves** 638

34-1 Evidence for Matter Waves 34-2 When Is the Wave Nature of Matter Important? 34-3 Light and Matter 34-4 Standing Waves 34-5 A Particle in a "Box"

chapter 35 **Atoms, Molecules, and Nuclei** 650

35-1 Hydrogenlike Atoms: Energy Levels and Size 35-2 The Helium Ion and the Helium Atom 35-3 The Lithium Atom 35-4 Absorption Spectra 35-5 Ionization Energies 35-6 The Size of Atoms 35-7 Electron Shells 35-8 The Pauli Principle 35-9 Chemical Binding 35-10 The Deuteron 35-11 The Nuclear Force 35-12 Complex Nuclei

chapter 36 **Change in Atoms and Nuclei** 672

36-1 The Conservation of Energy 36-2 The Conservation of Momentum and of Angular Momentum 36-3 Other Conservation Laws 36-4 Fission 36-5 A Closer Look at Alpha Decay 36-6 Classical Wave Models. Lifetimes and Barrier Penetration 36-7 Alpha Decay and Barrier Penetration 36-8 Lifetime and Energy Spread 36-9 The Emission of Photons Appendix: The Effect of Leakage on the Sharpness of Energy Levels: An Analogue Demonstration

APPENDICES 694

Appendix 1 **Table of Trigonometric Functions**

Appendix 2 **Table of Physical Constants and Conversion Factors**

Appendix 3 **Table of Elements**

ANSWERS TO SHORT QUESTIONS 700

INDEX 710

COLLEGE PHYSICS

INTRODUCTION

Man once feared the "illness of the sun," when the sun disappeared and the earth darkened. Then he learned about the complex motion of the moon. Eclipses became far easier to predict than tomorrow's weather. The moon has been circling our planet since long before the first dinosaur walked the earth. A man-made satellite can circle the globe for a long period without propeller, jet, or wings—a tiny synthetic moon. How do satellites move? How can we design our own? How can we visit the moon?

Physics enables us to answer such questions. It gives us the power to predict and to design, to understand and to adventure into the unknown. From what we learn in physics, new things are made. With new answers in physics, new questions are always arising. Many of these questions would never have been asked if physics itself had not been put to use.

Physics requires tools—tools of every kind. As with nearly all activities of human beings, the key tool of the physicist is his mind. Next, he needs language to make clear to himself and to others what he thinks and has done, and what he wants to do. Mathematics, which can be thought of as a special, supremely clear and flexible international language of relation and quantity, is also an important tool in his kit—and his own eyes, ears, and hands are very important indeed. He regards these as the first instruments for collecting information about the events of the world, a world which he tries to understand and to control. Then, to aid his senses, and to produce the special circumstances he sometimes wants to study, he must make use of a rich variety of other tools, instruments, machines, or contrivances.

Before the time of Galileo there were no astronomical telescopes. Once Galileo had put together two lenses to make an astronomical telescope and had discovered four moons revolving around Jupiter, more and better telescopes were designed and made. With their aid, new heavenly bodies were found, such as the many small planets called asteroids that move between the orbits of Jupiter and Mars.

New questions now arose. How could the complex motions of these moons and asteroids be explained? To answer questions like this, much of the special mathematical branch of physics called *mechanics* was developed. Beginning in the eighteenth century, rapid advances were made in this study of how objects move when subjected to complex forces. The new knowledge of mechanics led to the better design of machines. So we see that without the telescope, mechanics would have taken a slower course.

More recently, around seventy years ago, the beginnings of the understanding of atoms made it clear how to build much better air pumps than ever before. With these new pumps, a good vacuum was easy to obtain, and with a good vacuum experiments that formerly were impossible started researchers delving into the nature of electrons and atoms. This ability to produce a high vacuum, and the knowledge it brought, led to such varied practical achievements as radio and television tubes, concentrated orange juice, and atomic energy. This atomic research also clarified the very foundation of chemistry by making it possible to determine what holds two atoms together or what keeps them apart.

So the subject grows. It is like a great building under construction, not a finished structure around which you have only to take a guided tour. Though some parts are almost complete, and both useful and beautiful, others are only half done. Still others are barely planned. New parts will be started and completed by the men and women of your generation, possibly by you or your classmates. Once in a while, a finished room in this structure known as physics is found unsafe, or no longer large enough for new discoveries, and the room is abandoned or rebuilt. But the great foundations are well laid and stand on fairly solid ground. These remain unchanged though changes go on above them. It is the intent of this book to let you see the plan of the building, to show you what the builders have done, to look at some of the parts they are working on now, and to notice occasionally where the design is still incomplete.

1

Time, Space, and Matter

THE MOST universal instruments of physics are built into our bodies. Through our eyes we get most of our information about the world. Nearly as important are our ears, which bring us sounds. Then there are the various impressions we get through touch. These include the very delicate touch of our fingertips on which we rely for determining texture, the muscular senses of pushing and pulling, by which we form the impressions of weight and solidity, the feel of hot and cold, and our sense of balance. Smell and taste—more closely related to chemistry than to physics—are also important sources of our knowledge of the outside world. These tools, which continually bring information to us, are called the *senses.* Moreover, we do not merely sense the world passively; with our hands and backs and legs we handle and move some parts of it.

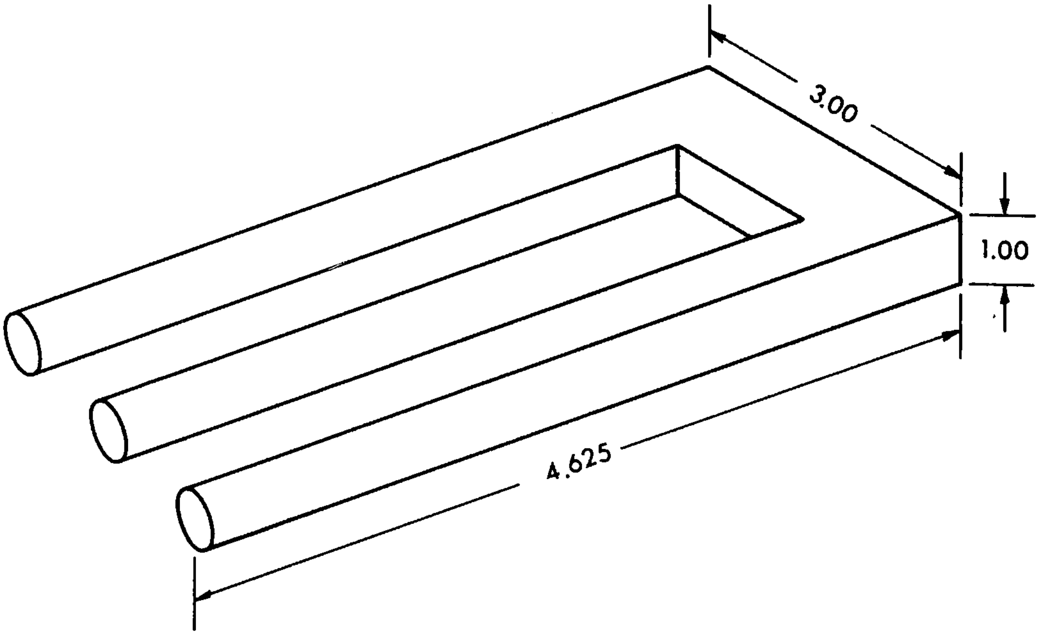
The five senses do not complete the list of tools that nature has provided us. We have in our brains and nervous systems a built-in set of instruments much more complicated and more highly perfected. Finding out how they perform is an unfinished part of science. Some idea of what these extra instruments do for us is gained by watching the behavior of simpler animals than ourselves, who may lack entirely something that we take for granted. An interesting example is the common octopus. This eight-tentacled monster is in fact a colorful, graceful, and clever beast. It is easy to set him a task by presenting him with some small moving object which he will attack. If he is rewarded with a morsel of food when he is right and punished by a small electric shock when he is wrong, he will soon learn to attack only some targets and not others. A blinded octopus uses his suckers to feel out the world, and will very quickly learn to tell a smooth rod from a rough one by feel, and even to tell a cube from a ball of the same diameter. But he cannot learn at all to tell a light rod from one ten times as heavy in the water but of the same outside size and shape. Even though you can see his muscles straining to hold the heavy rod, his brain never learns how heavy it is. He seems to have no way to tell where his eight legs are, or how they are bent, or by what paths they have moved. Lacking such internal instruments, he normally learns his way around by sight. The sensing of the positions and the strains in our fingers and hands and arms is a key part of our most complicated skills, like playing the violin or playing tennis.

The senses can be deceived. Optical illusions are familiar. Perhaps the most common one forms the basis of the movies. If you examine a strip of movie film you will see that it bears a sequence of slightly differing still pictures. Run them fast enough, and the eye will blend them into a smooth succession, which we recognize as motion. Your sense of temperature also can be fooled: if you hold one hand in a pan of hot water and the other in cold water, and then put both hands simultaneously into a pan of lukewarm water, the "cold" hand will feel hot, and the "hot" hand will feel cold. (If you haven't tried this, try it.)

This hot-cold illusion is surely inborn, and it depends on the way the nerve endings and the brain are built. But some illusions are not so clearly inborn. They depend also on things we learn, on rules or regularities, either man-made or in the natural world. The familiar and puzzling "space fork" of Fig. 1 - 1 shows this kind of illusion. Here there is a sort of pun done in lines instead of with words. A line represents either a corner in a solid block, or the edge between a solid and the empty air. The same line used in both *senses* leads to the mysterious fork. Placing a finger across the center of the three prongs kills the illusion, which depends on a mixture of the workings of sight and the rules of drawing.

Like the senses, all the other instruments of the physicist can be deceived—even the most accurate and sensitive, such as delicate balances, electronic meters, and timing devices. They all have their limitations. Testing the readings of his instruments, like questioning the first impressions on the senses, is part of the cross-checking which has to go into every conclusion made by the physicist. But this careful cross-checking gives him confidence in his instruments, just as our sense of touch helps us to confirm what we see with our eyes.

The physicist's instruments are especially subject to illusions like that of the space fork, in which the rules about the nature of the world, which he assumes to hold, play him false in different and unfamiliar applications. The reading of every instrument depends on a chain of understandings and is no surer than those understandings. Step by step the physicist has learned these pitfalls, and he struggles with new ones in every new experiment. By this time he has built up a wonderful and growing, though flawed, picture of the world. Behind that picture lie both the data he records and the theory with which he interprets the experiments. That theory and



1–1 An optical illusion. What parts of the drawing represent empty space and what parts represent solid material?

the data are intertwined and interconnected; both together make up physics.

1-1 TIME AND ITS MEASUREMENT

The quantities we measure in physics vary over a much wider range than those we deal with in our daily lives. Only within a limited range can we grasp them directly.

Table 1 gives us the order of magnitude of the time duration, expressed only to the nearest power of ten, of several things. Many of these intervals are somewhat variable—the human life span or the time for light to cross different rooms, for example. Consequently, it would not

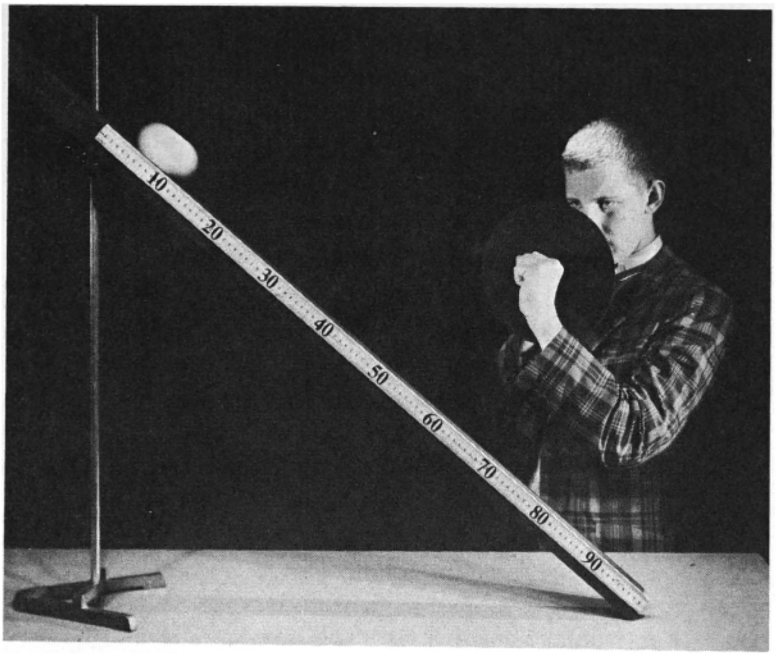
TABLE 1 *Orders of Magnitude of Times*

TIME INTERVAL IN SECONDS	ASSOCIATED EVENT	TIME INTERVAL IN SECONDS	ASSOCIATED EVENT
10^{18}	Expected total life of the sun as a normal star	10^{-1}	Time for bullet (30 caliber) to cover the length of a football field
10^{17}	Age of the oldest rocks Time elapsed since first fossil life Time elapsed since first land life	10^{-3}	Time for fly to beat its wings once
10^{16}	Time for the sun to revolve around the galaxy	10^{-4}	Time for one vibration of the highest-pitched audible sound
10^{15}	Time elapsed since dinosaurs	10^{-5}	Time during which firecracker is exploding
10^{13}	Time elapsed since earliest men	10^{-6}	Time for high-speed bullet to cross a letter of type
10^{11}	Time elapsed since earliest agriculture Time elapsed since earliest writing Time elapsed since the beginning of the Christian Era	10^{-7}	Time for electron beam to go from source to screen in TV tube
10^{9}	Human life span	10^{-8}	Time for light to cross a room
10^{8}	Time elapsed since you began school	10^{-9}	Time during which an atom emits visible light
10^{7}	Time for the earth to revolve around the sun (year)	10^{-11}	Time for light to penetrate windowpane
10^{6}	One month	10^{-12}	Time for air molecule to spin once
10^{5}	Time for the earth to rotate once on its axis (day)	10^{-15}	Time for electron to revolve around proton in hydrogen atom
10^{3}	Time for light from the sun to reach the earth	10^{-20}	Time for innermost electron to revolve around nucleus in heaviest atom
10^{2}	One minute		
10^{0}	Time between heartbeats (1 second)	10^{-23}	Time for light to cross the nucleus

make sense to specify them precisely. Giving just their orders of magnitude, as we have done in the table, is appropriate. Furthermore, the table shows an enormous range of time intervals, the longest a billion billion seconds and the shortest less than a billionth of a billionth of a second. To compare such widely different intervals it is convenient to compare orders of magnitude. For example, we see from the table that a day is eight orders of magnitude greater than (that is, 10^8 times as long as) the time for a fly to make one wing beat.

Notice that our own personal experience of time covers about ten orders of magnitude, from about 10^{-1} second, the smallest interval we can comprehend directly, to some 10^9 seconds, our life span. Another two orders of magnitude to 10^{11} seconds includes all we can learn from history, which is only a few hundred generations old.

To measure human life spans or the longer span of historical time, we have only to count the number of times the earth circles the sun. This human beings have been able to do since prehistoric times by observing the regular cycle of the seasons. We can use the rising and the setting of the sun—the rotation of the earth—to give us a smaller unit of time, the day. The day, subdivided into hours, minutes, and seconds, is commonly measured with ordinary clocks. Time intervals from tenths of seconds to a few tens of minutes we can easily measure with a stop watch. Beyond these time intervals, to measure both very long and very short times, we must devise special indirect methods.



1–2 Using a hand stroboscope. With six slits you can measure a time interval that is only one sixth of the time for one rotation of the disc.

The age of rocks, fossils, and the artifacts of early civilizations can be determined by using the laws of radioactive decay. Radioactive elements disintegrate into other elements at very well-defined rates that are different for different elements. From measurements made of the disintegration rate, we know that half of a sample of ordinary uranium, for example, decays into other elements in 4.5×10^9 years. Measurements of the ratio of the mass of uranium to the mass of the products of its decay in uranium-bearing rocks give the age of even the oldest rocks.

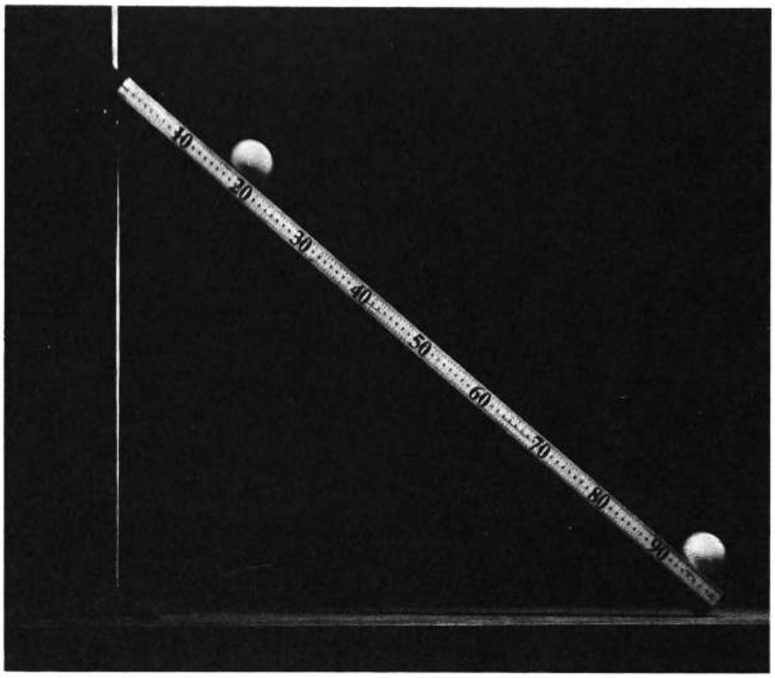
It is possible to make ordinary spring-driven stop watches that will measure shorter times than a tenth of a second, but it is not possible to start and stop such a watch by hand with enough precision to get meaningful measurements to less than that limit.

A simple device called a stroboscope can be used to measure time intervals considerably less than a tenth of a second. It consists of a large disc with slits spaced equally around the circumference. Figure 1 – 2 shows a simple form of stroboscope being used to observe the motion of a rolling ball. As the disc is rotated at constant speed, the slits pass before the eye and allow the operator to glimpse the falling object for a tiny fraction of a second at regularly spaced short time intervals. If the disc has three slits and is rotated once a second, the ball is glimpsed briefly every $\frac{1}{3}$ of a second, showing its successive positions at $\frac{1}{3}$-second intervals. If the number of slits is increased, the same rate of rotation shows the position of the ball at even shorter time intervals. For example, if a disc has 12 slits and makes one revolution per second, the ball will be glimpsed every 1/12 second.

Using such a simple stroboscope one can examine qualitatively the changes in position of the ball, but there is not time in each successive glimpse to measure its position; the brain does not work fast enough. If a camera is used in place of the eye, a permanent record of the successive positions of the ball can be obtained. All that is needed is a dark background, a white ball, and the camera set at "time exposure." To get a more even motion of the disc, an electric motor can be used to spin it. Figure 1 – 3 (a) is a strobe photograph of a rolling ball taken with a disc speed of five revolutions per second and with one slit in the disc. Figure 1 – 3 (b) was taken with the same disc speed of 5 rev/sec but with six slits in the disc.

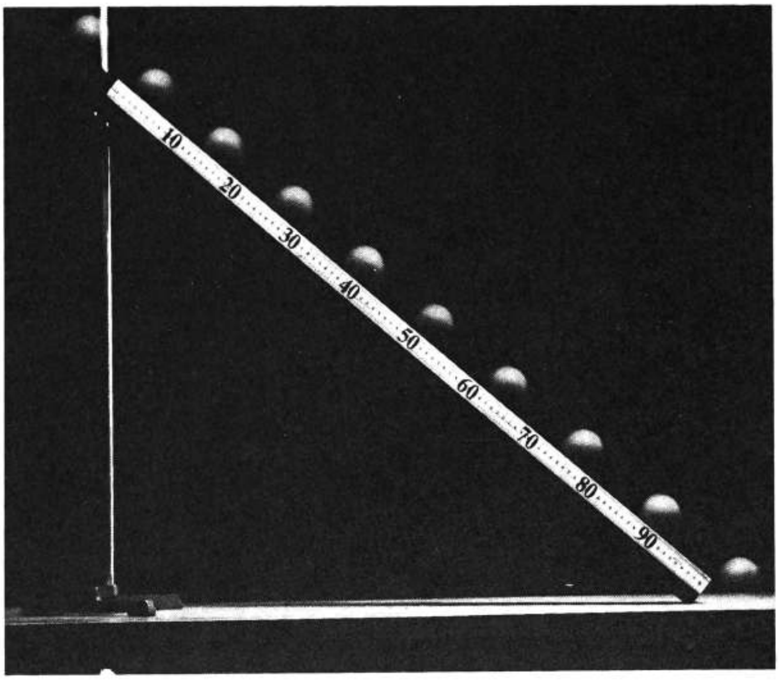
The stroboscope can also be used to measure the time of rotation of an object that is turning much faster than you can conveniently turn the stroboscope disc. As an example, suppose we use a stroboscope to measure the speed of rotation of a fan. We put a white mark on one of the blades and then observe the spinning blades through the stroboscope. If we spin the stroboscope disc fast enough, the marked blade will make just one complete revolution between glimpses and will always be seen at the same spot; the rapidly rotating fan will appear to be stopped. If our stroboscope disc has twelve slits and makes three revolutions per second, we get $12 \times 3 = 36$ glimpses each second and the time for one rotation of the fan is 1/36 second.

A stroboscope, like any other instrument, has its limitations. If the disc is spinning too fast, or if the slits are too numerous and too small, very little light can pass through a slit and you cannot see clearly. It is easy enough to measure the speed of rotation of a hand-rotated strobe disc. All we have to do is to count the number of turns we make in a minute. But what if we use an electric motor turning at high speed to drive the disc in order to measure very short time intervals? How can we know the speed of rotation of the disc? For many experiments, we do not need to know the disc's speed. So long as we are sure it is rotating at a nearly constant speed, the time interval between glimpses can be used as a unit of time. However, if we wish to measure this time interval in terms of a second we can use the method we have just described for measuring the speed of a rapidly spinning fan. We simply substitute the stroboscope of unknown speed for the spinning fan (it should have a mark on it whose position we can observe.) This stroboscope, once its speed of rotation is known, can be used to measure even shorter times, until we can go no further owing to the limitations of light intensity.



a.

1–3 Time-exposure photographs of the rolling ball shown in Fig. 1–2 taken through a stroboscope disc. In (a) only one slit was open and the strobe disc rotated at a speed of 5 revolutions per second. As you can see, a picture of the ball taken every 0.2 second does not give much information about its motion. It is impossible to tell if the ball is moving at constant speed, speeding up, or slowing down. In (b) six slits were open and the disc speed was again 5 revolutions per second, giving glimpses of the ball every 0.033 second. Now you can see that the ball is speeding up slightly as it rolls down the incline.

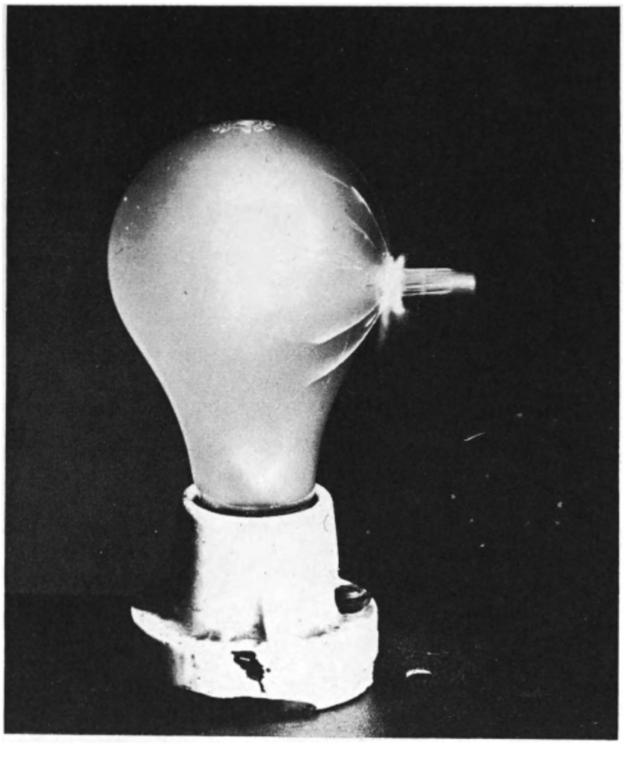


b.

To measure still shorter time intervals, beyond the limit of the simple rotating stroboscope disc, we must use another, similar technique—the high-speed flash—to stop a very fast motion.

A camera is set up in a darkened room; while the shutter is open, an intense flash of light of very short duration illuminates the subject. The camera film records the subject only while it is lit, so the resulting photograph shows an event which happened in an extremely brief time interval (Fig. 1 – 4). We can also use the flash technique to make a multiple exposure at equal time intervals on one piece of film just as we used a stroboscope to make Fig. 1 – 3. The photograph on the cover of this book is an example.

For very short time intervals, below direct perception, about three or four orders of magnitude to 10^{-5} second cover the ordinary motions and changes of the fastest-moving things men can make, even explosions. For intervals of still smaller orders of magnitude, measurements can be made with electronic oscillators, in which the much more responsive electrons act as the moving "parts" of the clock and measure time intervals as short as 10^{-10} second. Below that, all the way down to the shortest time intervals which are of physical interest—more than ten orders of magnitude beyond the reach even of electronics—we are in the realm of the atom. Here indirect reasoning alone gives us the time intervals that enable us to conjecture about what is going on in these tiny and swiftly changing parts of the world around us.



1–4 This high-speed photograph shows a .30-cal bullet as it smashes into a light bulb. The duration of the flash was less than one-millionth second. Note the rifling marks on the body of the bullet as well as the cracking of the glass. (Courtesy: Harold E. Edgerton.)

Just what makes a clock keep accurate time is a very hard and deep question. This is not simply a matter of the complicated works you see in an ordinary watch or clock. It is rather what you mean by time itself. Let us be satisfied with the idea that very carefully protected beating pendulums, or the newer electronic clocks which depend on the vibrations of a thin slab of quartz crystal, all count accurate time intervals. If these instruments are compared over periods of years and years, they agree with one another with high precision all over the world. Still newer ones, using certain actual atomic vibrations, are now being built. No one knows if there may not be some slight differences among all these means of measuring time. We only know that so far there have been none big enough to notice.

1–2 SPACE AND ITS MEASUREMENT

Time is one of the basic notions of science, and we have discussed it at some length. Perhaps you noticed that we were unable to confine the discussion to time and time alone. We were speaking also of positions and distances, of motion, of matter. These, too, are basic notions. Each of them is intertwined with all the others. It is impossible to deal with one without dealing with all the rest. To discuss them intelligibly, it is necessary to deal with them one after another, although they do not appear in nature one after another. They come together.

To use these notions skillfully, we must refine our understanding of each of them, and we must do this even though, strictly speaking, we have no place to start. The procedure—and all our history has shown it to be a very practical procedure—is to move back and forth among these basic notions, registering gains wherever we can and using these gains in turn to register further gains. We use our crude notions of space, for example, to refine our notion of time. Then we are able to use our refined knowledge of time to improve our notion of space. But, while we carry on this process, we must remember that physics itself is one indivisible subject dealing with the whole universe of which we are part. We subdivide it for our own convenience, but only so that later we may put it together again. What we can say with assurance is that size and distance, as divisions of space, determine the nature of the world just as much as does time.

The range of distances that we can measure covers, like the range of time, about 40 orders of magnitude, as Table 2 shows. In the remainder of this chapter we shall examine some of the methods we use to measure both large and small distances.

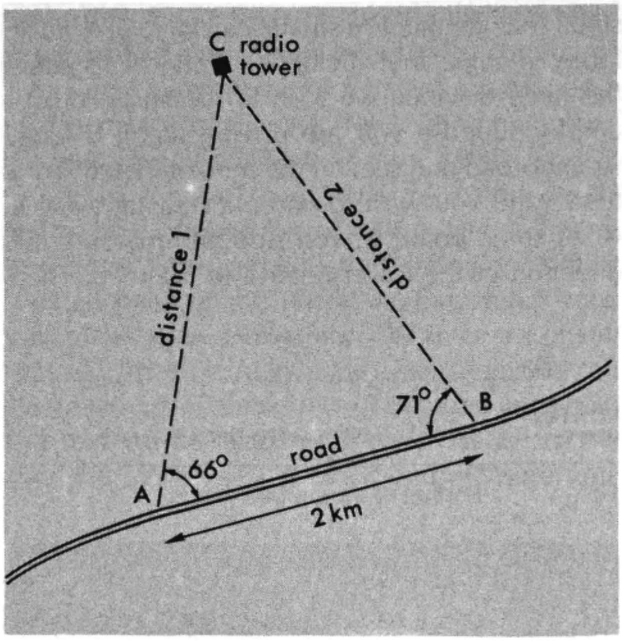
The method of laying standard lengths end to end can be used to measure quite large distances on the surface of the earth. The method is sometimes used in surveying, but often it becomes inconvenient. To measure the distance across a river, or the height of a mountain, or indeed the distance to a star, we can use a simple, indirect method. This method is based on the geometry of triangles and is called triangulation.

Essentially, this method consists of drawing a small triangle that is a small-scale map of a much larger triangle that includes as one of its sides the large distance we wish to measure. For example, suppose you are driving along a long, straight road and see, on the horizon, a tall radio tower and you wish to find out how far away it is. At some point, A, you stop and measure the direction of the tower relative to the road. Then

TABLE 2 *Orders of Magnitude of Distances*

LENGTH IN METERS	ASSOCIATED DISTANCE	LENGTH IN METERS	ASSOCIATED DISTANCE
10^{25}	Distance to farthest photographed object (a galaxy)	10^{6}	Radius of the moon
10^{22}	Distance to the Great Nebula in Andromeda (nearest galaxy)	10^{2}	Length of football field
10^{20}	Distance of the sun from the center of our galaxy	10^{-3}	Thickness of a windowpane
10^{19}	Distance to the North Star (Polaris)	10^{-4}	Thickness of a piece of paper
10^{18}	Greatest distance measurable by parallax	10^{-5}	Diameter of red blood cell
10^{17}	Distance to nearest star	10^{-6}	Average distance between successive collisions (mean free path) of molecules in the air of a room
10^{13}	Distance of Neptune from the sun	10^{-7}	Thickness of thinnest soap bubble still showing colors
10^{11}	Distance of Earth from the sun	10^{-9}	Size of molecule of oil
10^{10}	Distance of Mercury from the sun	10^{-10}	Average distance between atoms of a crystalline solid
10^{9}	Radius of the sun	10^{-12}	Average distance between atoms packed in center of densest stars
10^{8}	Mean distance from Earth to the moon Diameter of Jupiter	10^{-14}	Size of largest atomic nucleus
10^{7}	Radius of Earth Air distance from Paris to New York	10^{-15}	Diameter of proton

you drive, say, 2 kilometers along the road to a point B, stop, and measure the direction of the tower again. Then you can draw a map like that shown in Fig. 1 – 5 which is drawn to a scale in which one centimeter represents 0.5 kilometer. You can see, by measuring the distance AC in the figure, that the radio tower is about 2.7 kilometers from point A.



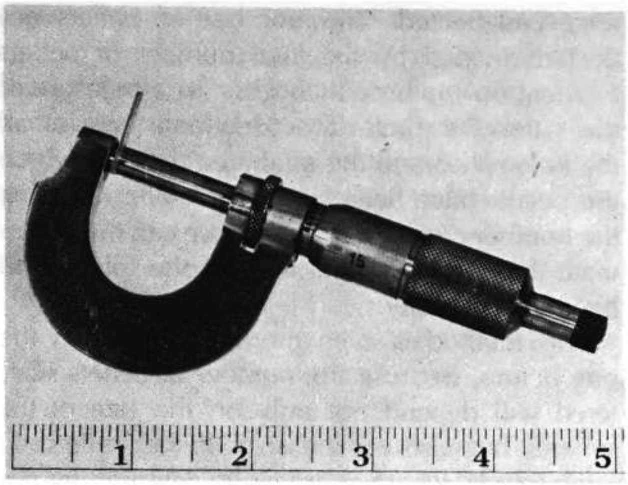
1–5 If the angles at A and B and the distance AB are known, the distance AC can be found.

An instrument can be constructed that will measure the angles of such a triangle and work out the answer automatically. An example of such a triangulating instrument, for measuring distances by simple sighting, is the range finder found in almost all good cameras. The base line (corresponding to the distance AB in Fig. 1 – 5), of the range finder can be no longer than the width of the camera, and distant objects thus appear at almost the same angle from both ends of it. Just where the sight lines cross is then difficult to determine. The camera range finder, therefore, measures only the distances to nearby objects. You can make range finders with longer base lines, and you will find that the longer you make the base line, the more precisely you can measure the angles to a distant object and the greater is the distance you can measure accurately.

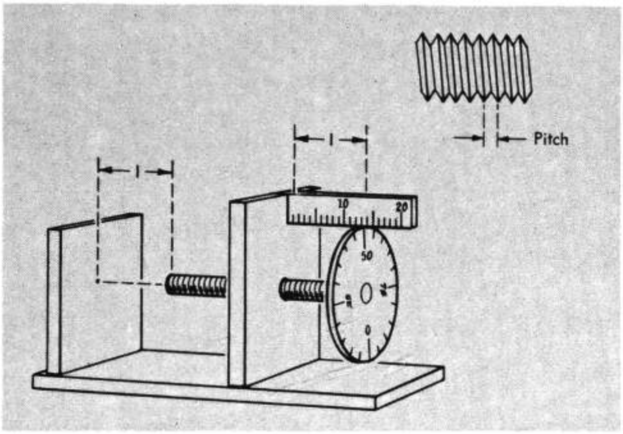
To measure the distances to planets, astronomers use base lines extending over half the earth. To measure the distance to the nearest stars we can use the diameter of the earth's orbit, the distance from one point on its path around the sun to the point reached a half year later. This sets the limit for measuring big distances geometrically. The distances to only a few hundred stars are small enough to be measured in this way. For greater distances in the universe, we must use other indirect methods.

If we go in the other direction toward the very small, we can still use straightforward geometrical methods. It is not hard to measure the thickness of a thin sheet of paper, if you have many of them. Stack up a hundred sheets and use a ruler to measure the stack; then you read on the ruler a hundred times the thickness of one sheet. This obvious indirect method is similar to what is often done in physics. Of course, if the sheets of paper are very different in thickness, the result will not refer to any real sheet but to an average of the thicknesses present. For many purposes this is good enough. What we obtain is the thickness of a sheet, assuming them all to be alike.

This sheet-thickness example shows how we can extend the basic idea of counting or spacing off to small distances. Another extension of counting to small distances is found in the use of a screw thread which can be used as a geometric amplifier of small distances. If a screw is turned through one revolution in a fixed nut, it advances only by the distance between successive threads, the pitch of the screw. By dividing the turn into, say, 50 parts, you can divide the advance of the screw into a 50 equal parts as well. This is the basis of the machinist's micrometer (Fig. 1 – 6 (a)). You can see from Fig. 1 – 6 (b) how such a device amplifies small distances. If the pitch of the screw in the figure is 0.1 millimeter, rotating the marked dial $\frac{5}{50}$ of a revolution would advance the end screw only 0.01 millimeter. This is a distance too small to measure directly, yet the screw thread amplifies it to one whole division as shown on the dial of the simplified micrometer. What we have here is a sort of lever that accurately amplifies a small distance into a larger, easily measurable distance. Other, similar amplifiers, some using a beam of light as the amplifying lever, help us to measure even smaller



1–6a A micrometer caliper. Very small lengths, such as the thickness of a piece of paper, can be measured with this instrument.

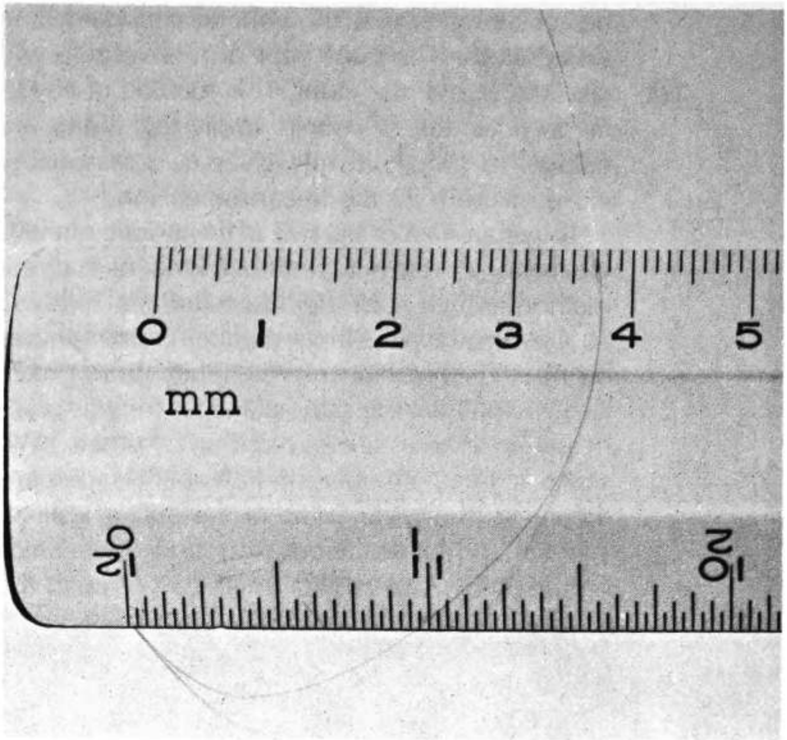


1–6b The basic part of the micrometer caliper, as shown in the simplified version sketched here, is a screw. The disc corresponds to the scale around the barrel in Fig. 1 – 6 (a), which enables us to measure a fraction of a turn.

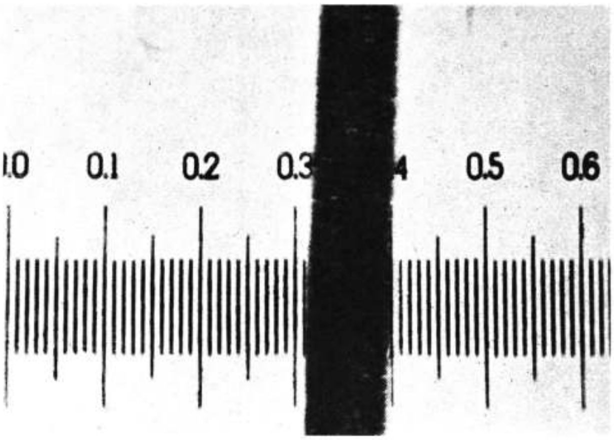
distances than can be measured with a micrometer.

To go further toward the very small we need to use still better amplifying devices, of which the most familiar is the microscope. With it we can see small objects, and measure them by placing tiny "rulers" right beside them (Fig. 1 – 7).

Below the range of microscopes we have to resort to indirect methods to determine sizes. Consider, for example, the determination of the size of an atom. We can imagine that, in a solid, atoms are "touching" one another, because solids are almost incompressible. If we know the number of atoms in a solid of given volume (Section 1 – 3), then that volume divided by the num-



1–7a The size of a human hair. In this photograph a hair has been placed across a millimeter scale. It is barely visible. How many hairs would have to be placed side by side to fill the space between two adjacent millimeter marks on the scale?

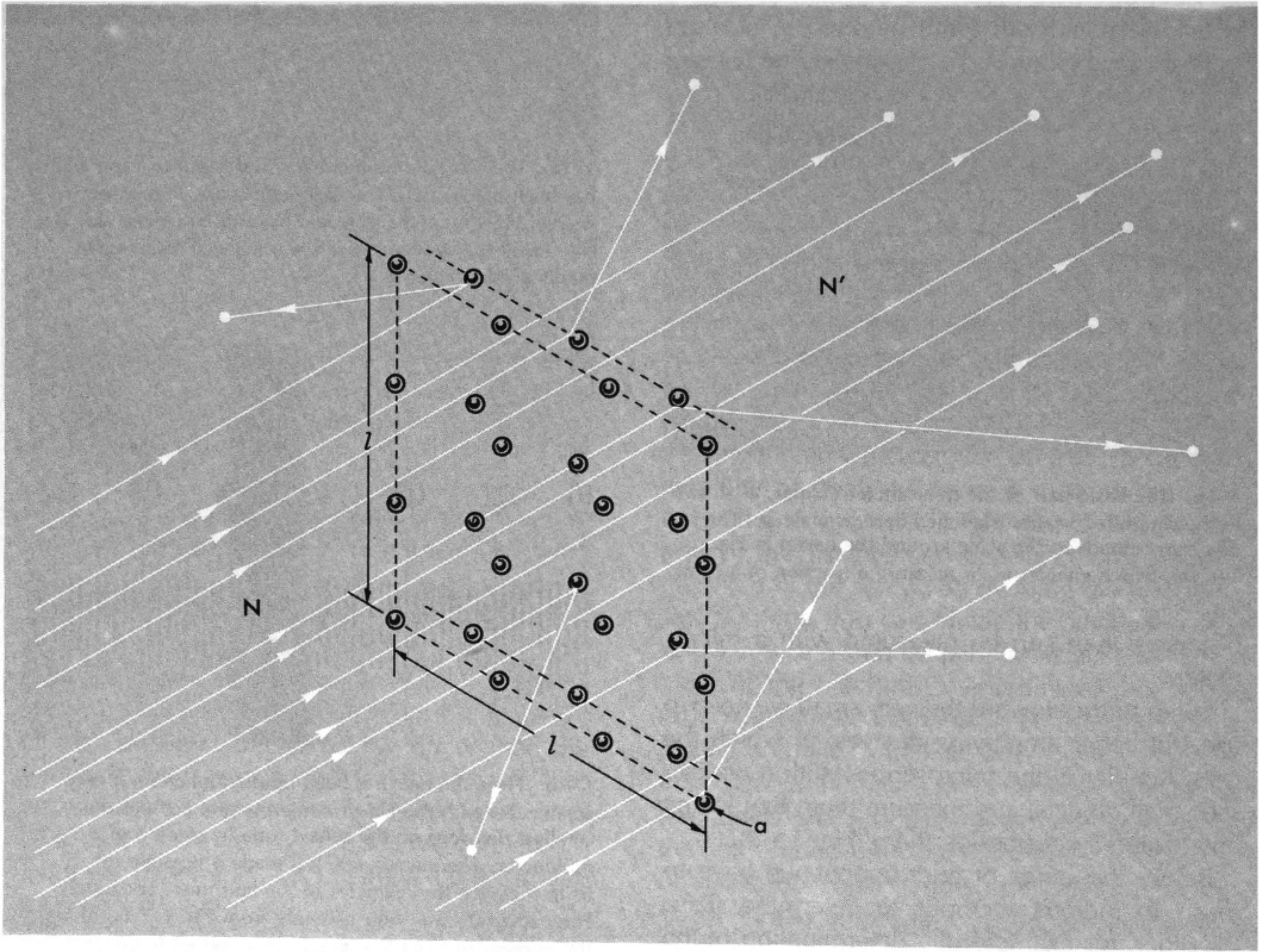


1–7b Here the hair has been photographed on a very small ruler with the aid of a microscope. Each of the smallest divisions on the ruler is one hundredth of a millimeter. The microscope has made it possible for us to measure the diameter of the hair more accurately. How accurate was your estimate from Fig. 1–7 (a) of the number of hairs that would be required to cover a millimeter?

ber of atoms equals the volume occupied by a single atom. The cube root of this volume will give the size of the atom. This method of calculation does not, of course, imply that atoms are cubical in shape; it only gives us a reasonably close measure for the linear dimension.

To get an idea of the size of the atomic nucleus we have to resort to another kind of indirect method which can be illustrated as follows: Suppose you have a large number of tiny spheres distributed over a given area. (They do not have to be actually in the same plane, as long as practically no spheres are covered up by others.) We shoot a large number of little pellets perpendicular to the area. Some of the pellets will hit spheres and be scattered away from their original direction, others will go straight through the area undeflected. The number of pellets deflected divided by the total number of pellets incident on the area then gives us a measure of the ratio of the total cross-sectional area of all the little spheres to the total area through which the pellets have been shot (Fig. 1 – 8). Knowing the number of spheres present, we can then calculate the cross-sectional area of one sphere and hence its diameter.

This method is even more indirect than the one before, because the number of pellets scattered will depend not only on the size of the spheres but also on the size of the pellets. In other words, what we are really determining is a sort of combined size of pellet and sphere, not really the size of an object as we know, for example, the size of a tennis ball. The idea that



1–8 The number of pellets deflected divided by the total number of pellets incident on the area $\left(\frac{N - N'}{N}\right)$, gives the ratio of the total cross-sectional area of all the small spheres to the area l^2. The cross-sectional area of a single sphere is therefore $a = \frac{(N - N')l^2}{Nn}$ where n is the number of spheres in the target sample.

objects have clearly measurable sizes independent of the way in which we measure them is characteristic of the macroscopic scale. In the microscopic world of nuclear dimensions sizes have a meaning only in relation to the way in which they are measured. It is in this light that we have to interpret the smallest dimensions shown in Table 2.

1–3 MATTER

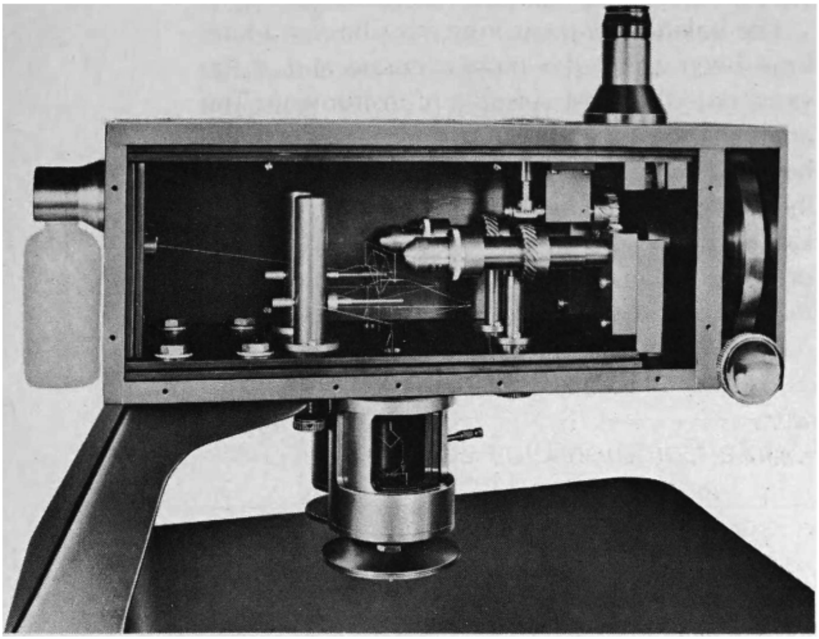
The third important notion in physics is substance, or material, — or *matter*, as it is more commonly called. It is one of the main successes of physics in our times that we have learned a great deal about the inner nature of matter. We have learned that all the differing materials — skin, bone, blood, rock, steel, nylon, air, even the sun — are composed of the same tiny building blocks, the atoms. Their combinations "spell out" the nature of the complex world in which we live, and even the nature of our bodies. Just as the twenty-six letters of the alphabet make up all the books that have ever been written in the English language, so the combination of a few building blocks makes up all matter with its great variety. For all the infinite variety of matter, we find that mass is a universal measure.

The balance for measuring mass has for a long time been about the most accurate and at the same time the most versatile of instruments. The analytical balance will compare masses of a few hundred grams showing differences between them as small as one part in a million. Table 3 is a list of a few common objects and their masses, presented to give you some idea of the range of masses that have been determined. The best

TABLE 3 *Orders of Magnitude of Masses of Some Common Objects*

MASSES FOUND ON THE BALANCE		MASSES FOUND BY OTHER MEANS	
MASS IN GRAMS	OBJECT	MASS IN GRAMS	OBJECT
10^{-6}	Check mark written in pencil	10^{-27}	Electron
		10^{-22}	Molecule of oxygen
10^{-4}	Fly wing	10^{-19}	Molecule of egg-white protein
10^{-2}	Postage stamp		
10^{1}	Coin	10^{-10}	Red blood cell
10^{2}	Paperback book	10^{-8}	Bacterial cell
10^{3}	One liter of water		
10^{5}	A man	10^{8}	House
10^{6}	One metric ton An automobile	10^{10}	Ship
		10^{18}	Mont Blanc
		10^{26}	The moon
		10^{28}	The earth
		10^{33}	The sun

laboratory microbalances, (Fig. 1 – 9) will measure differences as little as 10^{-10} kilogram, though only for postage-stamp loads. Large balances, rigidly built, sometimes using more complicated parts than the simple equal-arm device, give excellent results for masses up to almost a metric ton.



1–9 An ultra-microbalance. Even the delicate analytical balance is not sufficiently sensitive for some purposes. The quartz-fiber balance pictured above, however, is capable of measuring a mass to within 5×10^{-11} kg, although its load is limited to 2×10^{-4} kg.

Some objects, like the earth, are too large to put on any balance, yet plainly they possess mass. We can measure such masses in an indirect way. For example, if we know the volume of the object, and its ratio of mass to volume—its density—we can compute the total mass.

On the surface of the earth we can find a wide variety of materials of differing compositions and densities; in some places, such as the Grand Canyon of the Colorado River we can look down into the earth a few kilometers and we still find rocks similar to those on the surface. If we rashly assume that the deep, high-pressure interior of the earth has the same density as these exposed rocks, about 3×10^3 kg/m^3, it is not hard to estimate the mass of the earth, using the known radius, 6×10^6 meters, to compute the volume. The answer we obtain is about 3×10^{24} kg. it is interesting to compare this estimate with the value of 5.98×10^{24} kg given in a later chapter. The latter value is very much more precise; it rests upon yet another indirect method, a laboratory measurement of gravitation which you will study later in this course. Comparison of the two values demonstrates that the unseen interior of our earth consists of denser stuff than the familiar rocks of the surface.

Again, we must use indirect methods to measure extremely small masses. Some of these methods make use of electric and magnetic forces, but one method, which is very simple in principle, can be used to find the mass of an atom of a radioactive substance. The mass of a sample of the substance large enough to put on a balance is measured, and the number of atoms in the sample is found by counting the number of disintegrations with a radiation counter as the sample decays. The mass of a single atom is then found by dividing the mass of the sample by the number of atoms it contains.

1–4 ON THE LIMITATIONS OF MEASURING

We have founded all our measurements on one simple scheme. To measure the size of some physical quantity—say length, time, or mass—you first choose a unit; any length or time or mass you wish will do. Then to measure an interval larger than the unit, just "lay off" the unit as many times as it will go into the required interval. This is about what we naturally do with a ruler, a clock, or a balance. For anything left over after the count, or for any amount we want to measure which is smaller than the unit, we simply divide the unit into smaller equal parts—subunits you may call them—and take as many of them as we need to match the given magnitude. We measure a box and find its length to be 20 cm and something left over. Dividing our centimeter unit into tenths, we find that the part left over contains three of these subunits; so we say the box is 20.3 cm long. It is not hard to see that this method will work for any length that we want to measure. We can make the divisions finer and finer until irregularities in the edge of the object we measure, or in the markings on our ruler, limit the fineness of our measurements of its length.

Some measurements are not subject to the process of making smaller and smaller subdivisions for greater and greater accuracy. The counting of the number of people in a room, for example, has a natural unit, the individual. Here the whole question of smaller and smaller subdivisions is irrelevant. Some physical quantities do seem to have natural units. But we do not know whether space and time have such natural units: we only know that we have not yet found them. Until we find such units (if we ever do), we shall freely use any subdivision of our arbitrary unit of measurement to represent time and space.

We have just looked at the problems involved in the basic method of measurement by counting. In many real measurements a second type of problem arises. A measurement that is made by an indirect method is always based on special assumptions. In measuring the thickness of a piece of paper, for example, we made an assumption that the thickness of the sheets of paper was uniform. The measurement of large distances by triangulation also involves an assumption—one we are pretty familiar with in everyday life. We assume that the line of sight—which is the line that light travels to get to the eye from the object—is a straight line. Only if this is true will our method of sight triangulation work. Commonly we check the straightness of a board by sighting along it. We seem to accept the straightness of the path of light. Of course, it can deceive us and often does. The heat shimmer you see above a hot radiator or a sun-warmed surface tells you that here are sight-paths which are not straight and are constantly changing. If we wish a reliable answer when measuring long distances by triangulation, we must avoid looking through heated, disturbed air. We cannot measure the distance to a star by this means on a night when the star is twinkling very much as a result of changing air currents from the warm surface of the earth. We want a clear, still night, with the star well up in the sky.

Another assumption involved in measuring very large distances by triangulation is that the axioms of geometry are correct. They cannot be taken for granted, however. All assumptions that we make in measuring must be tested. The results of geometry and the straight path of lines of sight have been well tested, largely by the success of the whole picture we can build up. But we must always be on the watch, especially when using indirect methods in measuring things far from everyday experience, to see if such traditional assumptions can still be relied upon.

Even the method of using standard lengths has its problems. In very precise land surveys, for example, the temperature of the steel tapes used is measured meter by meter in order to correct for expansion or contraction. Here, because we employ physical objects, the direct method must be carefully scrutinized.

Experiments, and the measurements which very often—though not always—are involved in experiments, are the basis of physics. But we have now seen that a good many such results are not "pure." In fact, you can go further. Whatever experiment you try, whatever instrument you use, there is in the background some assumption or some principle, without which you would not gain any knowledge from what you are doing. Even the naked eye is such an instrument, not less than the most elaborate spectrometer. The way we find out about the world is to mix doing something—the experiment or the measurement—with thinking about something, the theory which underlies the kind of act we carry out. This intimate mix is not to be unscrambled; it is the use of the right sort of mix that makes progress in physics.

1–5 FRACTIONAL ERROR

We have used numbers to represent physical quantities, like times and lengths. Sometimes the numbers are precisely known, e.g., the definition of the centimeter is $\frac{1}{100}$ of the standard meter. Sometimes the number is only very roughly known and rather indefinite, like the width of the Grand Canyon (about 10^4 meters) or the mass of Mont Blanc (about 10^{12} tons). These numbers are all useful, each in its own way. They represent counts, in these examples, of the number of length or mass units in the quantities involved. The counts can be precise and finicky, or rough-and-ready. If a measurement is made to more than just order-of-magnitude accuracy, it is very helpful to include with the statement of size some indication of the intended accuracy of the statement. One way to show this is to finish off the statement with a judgment about the range in which the number is expected to fall. For example, if you measure your height by having someone mark the place the top of your head comes when you stand at the wall, you might say your height is 178.1 centimeters, but you do not expect this to be exactly right—to give the same result to the nearest millimeter upon repeating the whole procedure. How you stand is too variable for that kind of precision. So you might record your height as 178.1 cm, but pretty surely not more than 178.6 or less than 177.6 cm. A brief and professional way to say all that is to state your height as 178.1 ± 0.5 cm (to be read out loud as "178.1 plus or minus 0.5 cm"). That would make clear what kind of precision such a measurement sought to achieve. If you used a micrometer to measure the diameter of a steel ball bearing, you might give the result as 2.143 ± 0.002 cm, if you felt that the mark might shift by one or two thousandths of a centimeter as you tried different measures at different diameters of the ball, with different judgments about when the measuring jaws were drawn up tight. It is pretty clear that doing arithmetic with such numbers requires a little thought. If you stand with the ball bearing balanced on the top of your head, you are not known to be 180.243 cm high, even if that is the sum of 178.1 cm and 2.143 cm. It would be better not to claim so much, and remember that the spread in measurements of total heights is still about 0.5 cm, for the tiny error of the micrometer is lost in the big uncertainty of the body height. The best one can do is to write 180.2 ± 0.5 cm. Keeping the plus-minus helps the reader to see what you have done, no matter how you state the results.

This spread is very often given as a percentage, or fraction. If a surveyor measures a distance with great care he might get 100.132 meters ± 0.3 cm. His work is a great deal more accurate than that done when the width of a book page is measured to the nearest millimeter with a ruler, even though his error is something like three times as big as what anyone would perhaps make on the page in ten seconds' work. This sometimes finds expression in another way when the estimated spread of measurements, the *tolerance,* is stated, using decimal fractions, or percentage. Thus the surveyor would say his length was 100.132 meters ± 0.003%, while the page is just 20.1 cm ± 0.5%. These numbers can be translated back into the other sort of plus-minus statements by doing the indicated calculation.

Such ways of expressing results will become more and more natural to you as you learn more about experimenting in physics. There is no technique in physics more important than the writing of physical quantities with the magnitude, the unit, and a reasonable estimate of the error all made clear.

1–6 BLACK BOXES

The careful reader will have noticed that we have already used some of the instruments of physics without thoroughly understanding their operation. We used the balance and the gravitational pull of the earth to determine mass, but we have yet to study gravity. We used telescopes and range finders, but we have yet to study light. We used clocks we cannot repair, without saying how they work. Such an optimistic and disorderly procedure is usual for physicists. It is probably necessary; no one has ever done without it. For physics is unlike mathematics: the more logical mathematician makes his world, by strict definition and proof, taking at most a few matters as undefined.*

The physicist is luckier in some ways and yet less fortunate in others than the mathematician. For the world is here, and the physicist does not have to invent it. He is forced to study it by the means at hand: his senses, his reason, and the instruments he can devise. Those instruments themselves have to be understood, if his work is to be complete. But then where can he start? For if he cannot talk of mass until he understands gravity, he will be a very quiet scientist indeed. Instead, he has learned the use of "black boxes." Let us first explain this curious phrase.

Suppose you looked at the radio rack in a commercial airliner. You would see one or a collection of black metal boxes. Various wires go in and out of the boxes, interconnecting them, or going out to the antenna or ground, to the power source, or to a point of "display" like a loudspeaker or a meter. If you peered under the box lid, you would see a maze of colored wires and little bits and pieces of electronic equipment. You would probably not know the purpose of each wire; yet you could turn the dials and make the equipment operate.

---

*It is one of the most striking results of modern mathematics (discovered by Gödel in 1932) that even the mathematician cannot proceed fully logically, proving every statement to be true or false; he must always remain incompletely logical, or stay within rather trivial limits.

Such experience has given us a useful and expressive phrase. We speak of a physical system of any kind as a "black box" when we make use of it without analyzing how it functions. With some care we can successfully use an instrument whose operation we do not understand.

The eye is a fine example of a black box. But we know from innumerable crosschecks how to use it, how to judge light and dark, or large and small, or fast and slow. We know its limitations, too. So far in this book, the microscope is a black box. Yet it is not likely that the use of the microscope is deceptive. To measure the size of a hair with it, we simply look at the hair against a finely ruled grating. The experience of ordinary viewing convinces us that when we see the hair beside the rule we can measure its size pretty well, even if we are looking through a microscope at the two of them. We are using the microscope as a black box; but we gain confidence in it by use.

Quite generally, we find out how to use black boxes by using them to study known objects, objects like the little ruler which we feel we can understand from earlier experience. For example, we can use the split-field range finder of a camera even when we don't know how it works. We learn by use that when we fit both parts of an object together in the viewer, the focus of the lens is automatically adjusted to give a clear photograph. The use of a black box in known physical situations allows us to learn how it performs even if we do not find out why. When its performance is known, we can use it to make new measurements. On the basis of use we can decide what display of the instrument (black box or not) corresponds to the value of the quantity we want to measure. This process, learning the performance of an instrument by measuring what we already know, is called *calibration*. Usually we calibrate measuring instruments, as you may have done for your microbalance or your range finder.

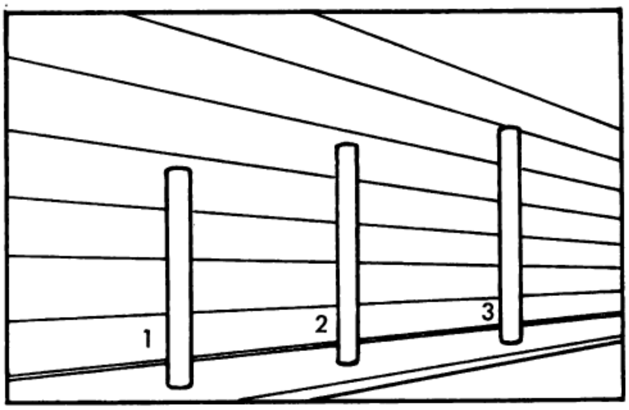
Physics is a great enterprise of human beings. No one knows all there is to know about physics or can do it all. All of us use some black boxes, and all instruments are, at least, partly black box to everyone. The principle of the equal-arm balance is clear, but generations of design and experiment lie behind the shape of the beams made so that they will not bend, the support of the pans made so that they hang vertically, and a thousand other points. For every user but a skilled and experienced designer of balances, some of this remains a black box. Even the designer probably regards as black box the agate bearing on which his knife edge rests. He knows agate is hard and durable, but he may not know why. The properties of the agate depend on its molecular structure; that may be a black box even for the designer.

Although we have used the microscope as a black box, it cannot be all black box to everyone. Most man-made things cannot be complete black boxes, for someone has to make them. They are partly black box because we are aware of so much that we have yet to understand.

Curiosity to "open" the black boxes is needed for the understanding of physics. But also one needs the judgment to know when and where black boxes may be safely used. Confidence in the black box comes with calibration, use, cross check, and ultimately with "lifting the lid" and tracing out the method of operation. What remains a black box today may be opened in a generation; but even its opening is likely to involve the skillful use of all sorts of black boxes, black boxes which we have yet to see.

QUESTIONS

1. Compare the apparent heights of columns 1, 2, and 3 in Fig. 1 – 10. What is the actual relationship of heights? Did you use a device to check up on your senses?



1–10 For Question 1.

2. A stroboscope is constructed by making 10 slits in an otherwise opaque disc. Find the number of revolutions per second of the stroboscope which are required to "stop" a loaded doorbell clapper which executes 10 back-and-forth swings per second.

3. A rotating disc with a radius marked on it is seen through a revolving strobe with one slit which is rotating at nearly the same rate as the disc. If the radius appears to move slowly backward, is the strobe rotating at a higher or lower rate than the disc?

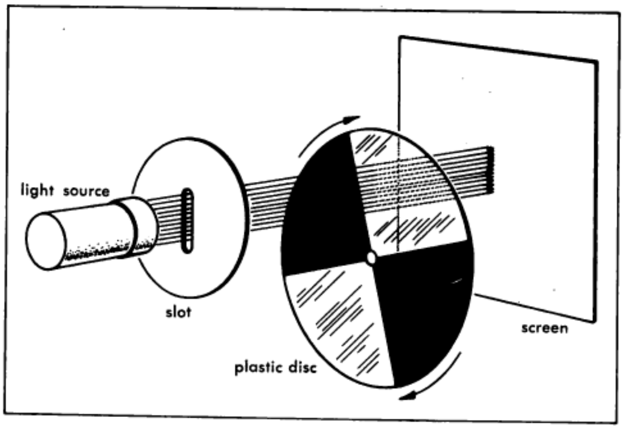
4.* What will you see if you look through a rotating stroboscope at its image in a mirror?

5. How many times per second should a light flash to show images 10 centimeters apart of a projectile traveling 1000 meters per second? (1 meter = 100 centimeters.)

6. In a certain experiment, a narrow beam of light is "chopped" into a series of pulses by a rotating transparent plastic disc with sections blackened as shown in Fig. 1 – 11. The disc is rotating 30 revolutions/second.

   (a) What is the time interval from the beginning to the end of one light pulse?

   (b) What is the time interval from the middle of one light pulse to the middle of the next?



1–11 For Question 6.

7. In time-lapse photography, plants and their blossoms may be seen to develop at a rate comparable to that of the motion of animals. If the total growth of a plant which would require 50 days is to be shown on a 10-minute film, what interval should elapse between the successive pictures? (Asssume that the film is to be shown at 24 frames/sec.)

8. A project for a homemade timepiece: Galileo performed some of the first important experiments with the motion of falling bodies before the days of accurate clocks. To measure the short time intervals involved in his experiments, he used a simple clock that you can make for yourself. In the bottom of a tin can, punch a small hole with a nail. Keep the can nearly full of water and measure the amount of water that falls out in 10, 20, and 30 seconds. In this way you can learn to read the clock in seconds.

   What are the major sources of error in this device? Can you reduce the errors? Remember that Galileo used a clock like this in discovering some of the major principles of physics. (See *Two New Sciences*, by Galileo.)

9.* From Table 2, calculate the order of magnitude of the ratio of the size of a molecule of oil to the size of the largest atomic nucleus.

10. Estimate the order of magnitude of the following items:

   (a) The number of drops of water in a bathtub full of water.

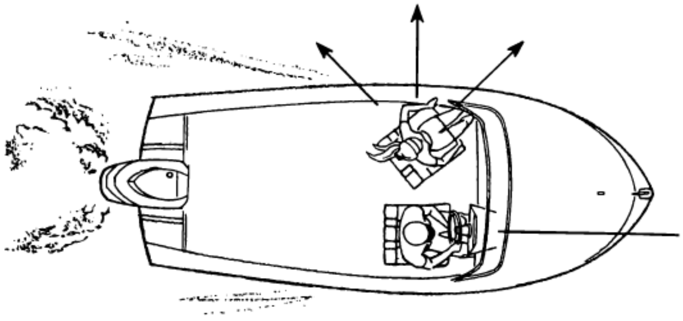
   (b) The thickness of a leaf of your textbook.

   (c) The area of the earth.

   (d) The volume of the earth.

11. Choose an arbitrary unit such as the length of the first joint of the thumb or the length of a fingernail. Mark off a scale, using this unit, along the edge of a piece of paper. Use this scale to measure the length and the width of this page. Determine the ratio of the length to the width. What are the units of this ratio? Compare the ratio that you obtained with the ratios that others obtained for this same object using different units. What is the effect of different units in measuring ratios?

12. The skipper of a small boat wants to determine his distance from points on shore without instruments other than his watch and the known speed of his boat. To do this he sets up sight lines at 45°



1–12 For Question 12.

and at 90° to the centerline of the boat, as shown in Fig. 1 - 12.

Suppose that you are the observer riding in this boat. The skipper is holding a clear, straight, and steady course at 6 meters per second. What observations and calculations must you make to determine your shortest distance to a house on shore?

13.* A student has found that when he holds a pencil vertically at arm's length, it just obscures his view of a distant radio tower. The distance from his eyes to the pencil is 60 cm. The length of the pencil is 15 cm, and he knows the tower is 75 m high. How far is he from the tower?

14. What is the diameter of the moon? Assuming that the moon is 3.8×10^5 km from the earth, you can determine its diameter by the following method. Place two strips of opaque tape 2 cm apart on a windowpane. Make a pinhole in a card and then observe the moon through the pinhole and between the two strips. Back away from the window until the moon just fills the space between the two strips. Measure the distance from the card to the window. Using the geometry of similar triangles, calculate the diameter of the moon. (Do *not* attempt to use this method to determine the diameter of the sun; the bright sun would be harmful to the eye.) Could this method be used to determine the size of a star?

15.* What is the diameter of the hair shown in Fig. 1 - 7?

16.* Which is larger, the mass ratio of the sun to the earth, or of an oxygen molecule to an electron?

17. If you have an equal-arm balance and a 1 - kg mass, describe how you would make 1/2 - kg masses.

18. Suppose you had a 1 - kg mass and you wanted to make another, but the only balance you had was an equal-arm balance that you did not quite trust. You put the 1 - kg mass on one pan of the balance and worked over the new mass until the beam of the balance stood horizontal when the new mass was placed on the other pan.

(a) How could you tell whether you had equal masses hanging from balance arms of equal length, or whether the balance arms were not quite equal in length and you had made the new mass just different enough to compensate?

(b) To finish the job, you might have to adjust the balance arms as well as the new mass. How could you tell which of the two masses was larger and which of the two balance arms was longer?

19. You probably know that the shape of the earth is a slightly flattened sphere. The polar axis is shorter than the diameter of the equator by about 40 km. Is the earth more nearly a perfect sphere than a basketball?

20. Check your time sense by trying to estimate when an interval of 1/2, 1, or 3 minutes has elapsed.

(a) Record your error in each case.

(b) Are the errors appreciably different fractions of the time intervals involved?

21. A student measures a block of wood and records the following results: length = 6.3 cm, width = 12.1 cm, and height = .84 cm.

(a) What is the volume of this block?

(b) Assume the length and width measurements to be correct; however, you can see that the height measurement may be off by 0.01 cm either way. How would this change your answer for the volume?

(c) What fraction is this of the total volume?

22. A bag contains 25 identical marbles. A student makes measurements on one of them and determines its volume to be 1.76 cm^3. What should his answer for the total volume of all the marbles be?

23. A person reports seeing a "flying saucer" 200 feet long, at an altitude of 6000 feet and flying at an estimated speed of 2000 miles per hour. Discuss the accuracy of this observation.

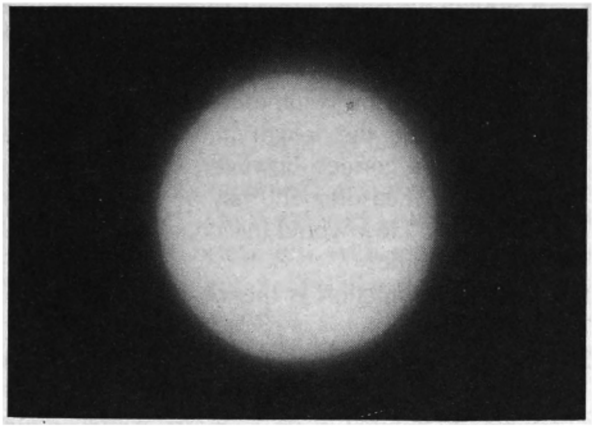
24. Make a pendulum from a piece of string and a small weight. Adjust the length until it takes one second to make a complete swing to and fro. How great is the error of your pendulum over a one-minute period? What fraction of the total time is the error?

25. (a) A simple range finder can be improvised by using two meter sticks to establish sight lines to the object. The amount these deviate from parallel position is a measure of the nearness of the object sighted. If the meter sticks form equal angles with their base line, and are separated by one meter at the near end and 96 centimeters at the end away from the observer, what is the distance to the sighted object?

(b) If the meter sticks are not sufficiently straight to give an accurate reading on the separations, an error in range estimation will result. What difference in the estimated range would result if the front measurement were 96.2 instead of 96.0 cm while the rear measurement remained at 1 meter? Express this error as a percent.

(c) As the sticks are made more nearly parallel, failure to sight the sticks exactly produces a much larger error of range estimation. Compare the ranges when the front ends of the sticks are 99.0 and 99.2 cm apart while the rear ends remain at one meter.

26. Figure 1 – 13 is an out-of-focus photograph of a circular plate. From the photograph, how precisely can you determine the diameter of the plate?



1–13 For Question 26.

27. Can you determine the diameter of the earth including its atmosphere?

28. If you hang a mass on the end of a long spring supported from the ceiling the mass will vibrate up and down. You can measure with a watch the number of vibrations it makes per minute. Suppose you do a series of experiments with different masses and different springs and arrive at the following conclusions:

In general, the same mass suspended from different springs vibrates up and down at different rates.

The larger the mass hung from a given spring, the lower the rate of vibration.

In particular, quadrupling the mass suspended from a given spring decreases the rate of vibration to very nearly 1/2 for heavy masses.

Quadrupling the mass reduces the rate of vibration very slightly for very light masses.

Can you explain all the above observations? If you cannot, the apparatus is a "black box."

(a) How could you use this black box to measure mass?

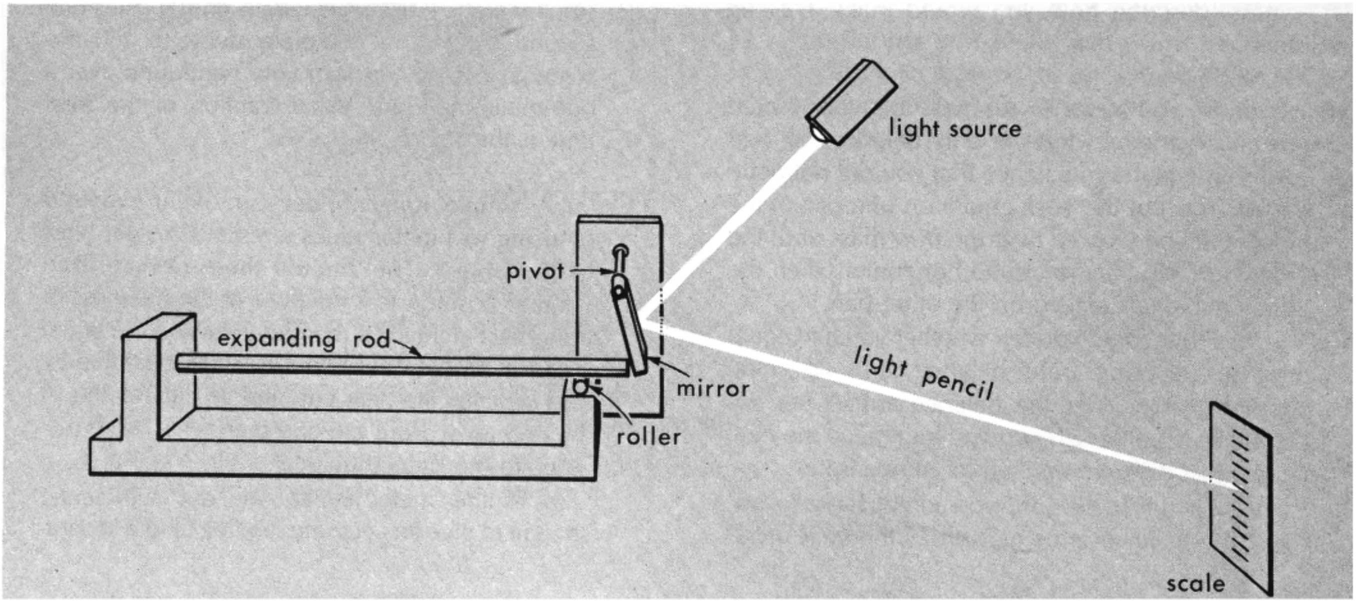
(b) What do you think would limit the range of masses that you could measure with this apparatus?

29. Suppose you make your own ammeter to measure electric current and you equip it with an arbitrary scale that reads from zero to ten. Compared with an accurate ammeter, the homemade one reads 2.3 scale divisions for a current of 3.0 amp, and 4.8 scale divisions for a current of 6.00 amp. What current is flowing when the homemade meter reads 3.5 scale divisions?

30. (a) How would you calibrate the apparatus in Fig. 1 – 14 to measure the expansion of the metal rod when it is heated?

(b) Could you calibrate the apparatus to use it as a thermometer?

1–14 For Question 30.



2

Data, Graphs, and Functions

IN THE PRECEDING chapter we have used geometry and simple arithmetic in discussing the measurement of time, space, and mass. In this chapter we shall see how useful tables and graphs are in displaying the data we accumulate in the laboratory, and how they can be employed to predict the results of experiments we have not yet tried. Furthermore, we shall see that graphs can be very useful in finding out what mathematical expression describes the relationship between two measurable quantities. Finally, we shall see how a knowledge of certain functions involving physical quantities can be used in predicting what will happen if we scale the physical dimensions of things up to very large sizes or down to very small sizes. We shall begin by discussing an experiment that you can do with the bell timer whose rate of vibration you measured in the laboratory with a stroboscope.

2–1 TABLES AND GRAPHS

In the laboratory experiment with the bell timer we might measure the number of vibrations per second of the clapper for various loads attached to the clapper. We should then have a collection of measurements relating these two quantities which can conveniently be presented in an ordered table of data in which we write the two columns of data in order of increasing load as shown in Table 1. A glance at such an ordered table allows us to find the value of one of these quantities for any value of the other quantity which appears in the table. However, suppose we wish to know the values for a case that does not appear in the table—for example, the frequency of the clapper with a load of 3 gm. What prediction can we make from the table? The table shows that over the entire range where measurements were made, the frequency decreases as the load increases. Thus, we expect the frequency with a load of 3 gm to be between 39 vibrations/sec and 35 vibrations/sec, the values corresponding to loads of 2 gm and 4 gm respectively, but it is hard to tell which intermediate value the frequency will have. A first approximation is to

TABLE 1

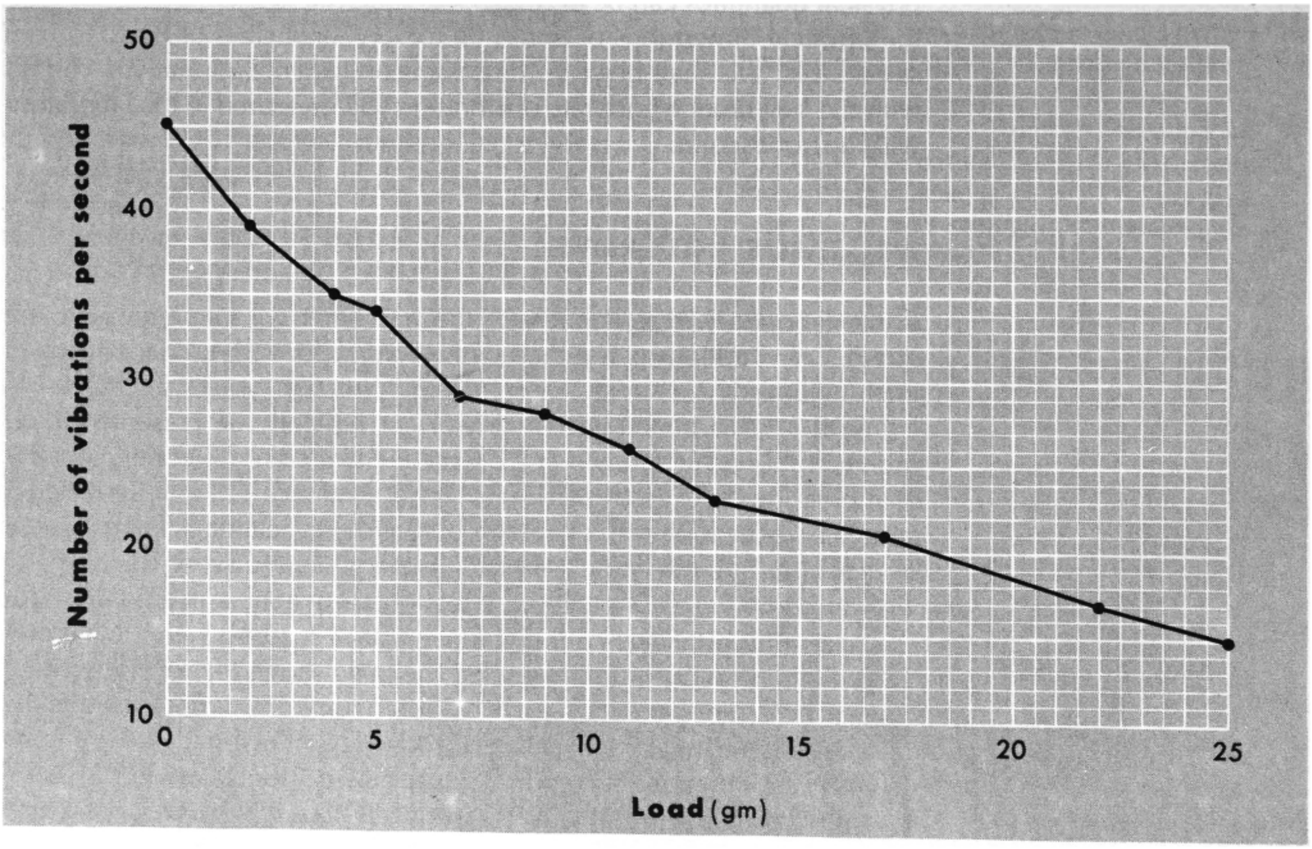
LOAD	FREQUENCY
(gm)	*(vibrations/sec)*
0	45
2	39
4	35
5	34
7	29
9	28
11	26
13	23
17	21
22	17
25	15

say that the frequency for 3 gm will be halfway between 39 and 35, that is, 37 vibrations/sec. In so doing, we make little use of the table as a whole. We determine the frequency of the clapper with 3 gm on the basis of data for 2 and 4 gm alone. This kind of interpolation between two measured values of a variable can be presented graphically: We plot the experimental points on graph paper and then connect adjacent points with straight line segments. From this graph we can then read off the frequency for any value of the load up to 25 gm [Fig. 2 – 1 (a)].

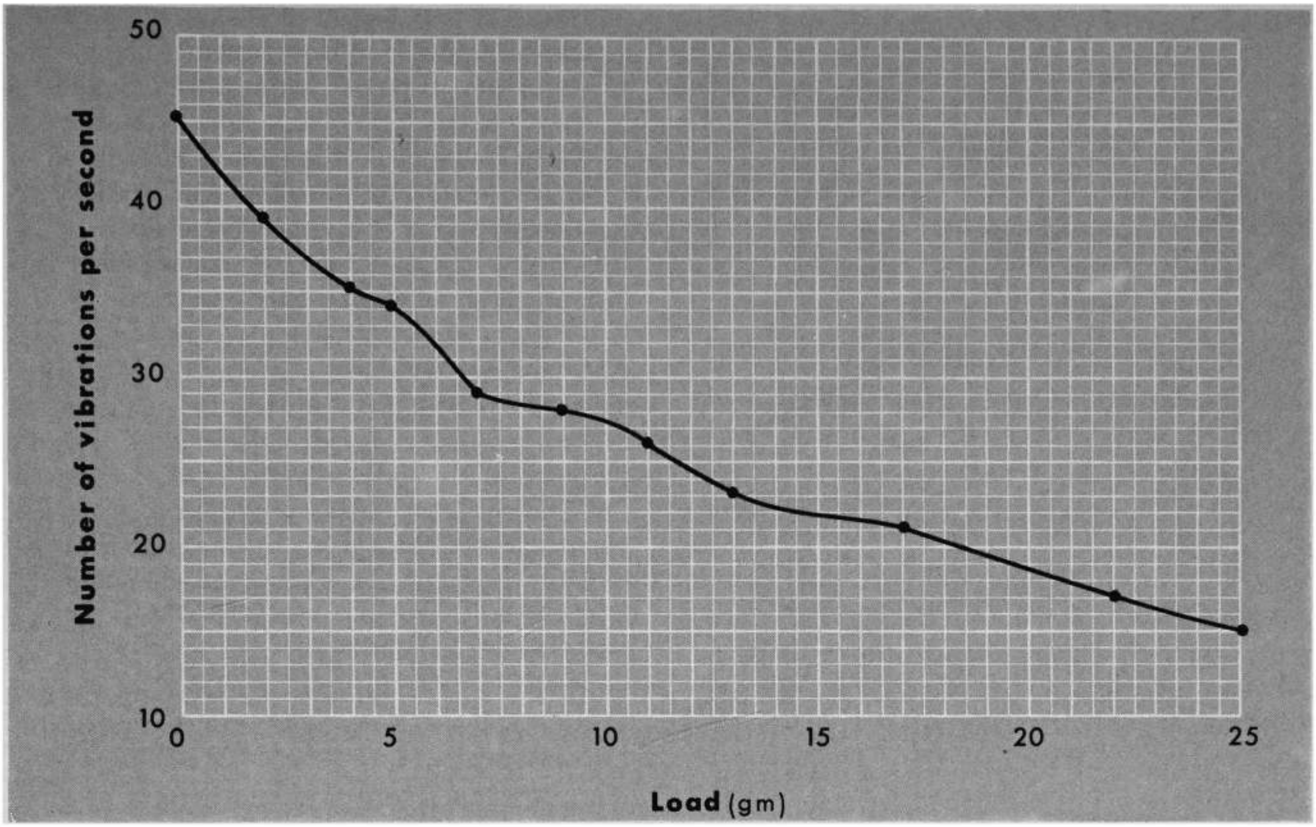
Such a simple presentation has shortcomings. First of all, notice that the graph has sharp corners at the experimental points. If we had made the measurements with loads different from those we actually used, these corners would be in different places from those in Fig. 2 – 1(a). Therefore, they cannot have any real significance. We could try to correct the situation by smoothing the corners [Fig. 2 – 1(b)], but this still gives each experimental point more importance than it deserves since again different loads would have given smoothed corners in different places.

The data in Table 1 are given to the nearest gram and the nearest vibration/sec. We can indicate this on the graph paper by drawing little rectangles, 1 gm wide and 1 vibration/sec high around the data points [Fig. 2 – 1(c)].

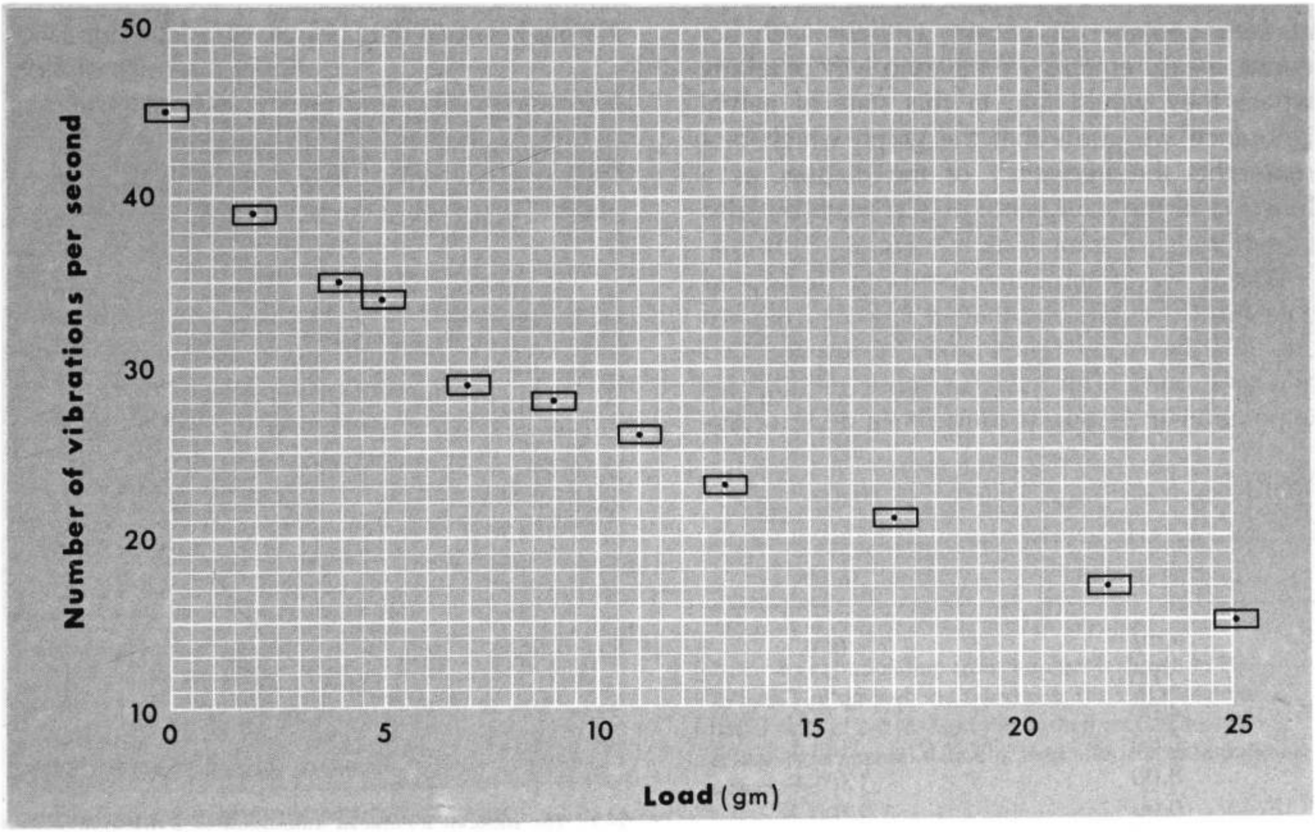
Experience has shown that if we had a more accurate table of data, we should find that most of the more accurately measured points would



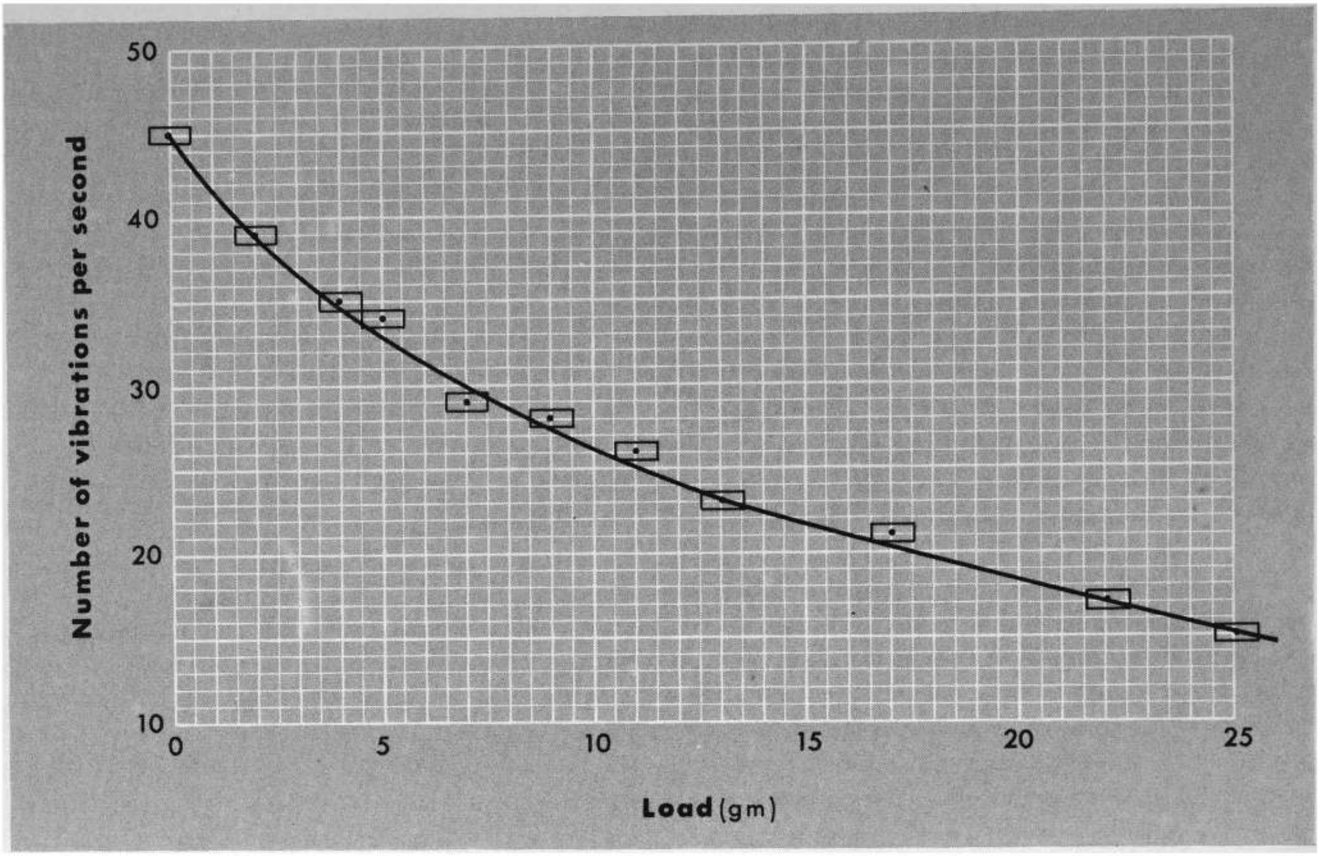
2–1a A graph made by connecting experimental points by straight segments.



2–1b The same graph as Fig. 2–1 (a), but with the sharp corners rounded off.



2–1c The experimental points and their uncertainties.



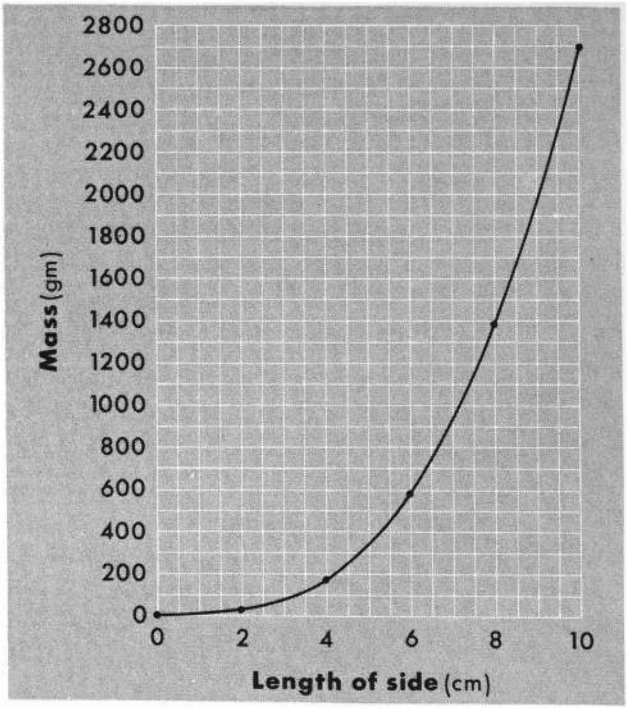
2–1d A smooth curve which summarizes the results of the experiment.

fall inside the little rectangles but not necessarily at their centers. Of course, without such accurate measurements we have no way of telling where they would fall. In this state of partial ignorance we guess that the graph which best represents the frequency of the clapper as a function of the load is not only smooth but also free of wiggles, something like Fig. 2 – 1(d).

When we interpolate between measured data from a curve like that in Fig. 2 – 1(d), we are utilizing many points and not just the two adjacent ones. This gives a more reliable interpolation than one made from a table of experimental data, particularly if the graph curves sharply between data points. Consider Fig. 2 – 2,

TABLE 2

LENGTH OF SIDE	MASS
(cm)	*(gm)*
2.00	21.6
4.00	173.0
6.00	583.1
8.00	1381.0
10.00	2700.2



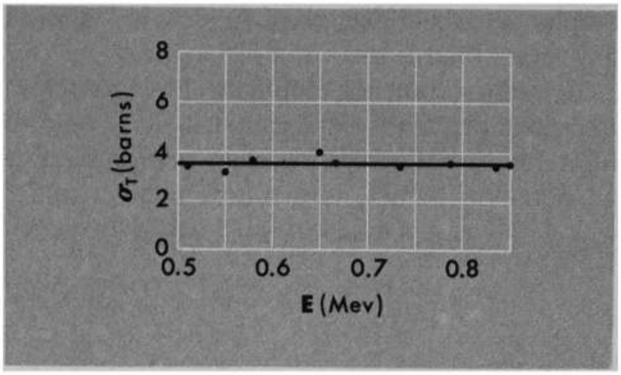
2–2 The mass of a cube of aluminum as a function of the length of one side.

for example, which is from data obtained from very precise measurements (Table 2) of cubes of aluminum of different sizes. Interpolation from the table alone predicts that a cube 3.00 cm on a side will have a mass of 21.6 gm + 0.5 × (173.0 − 21.6) gm = 97.3 ± 0.05 gm. But the graph, which shows the curvature between the data points, predicts a mass of 70 ± 5 gm, which is likely to be closer to a value found by a direct measurement.

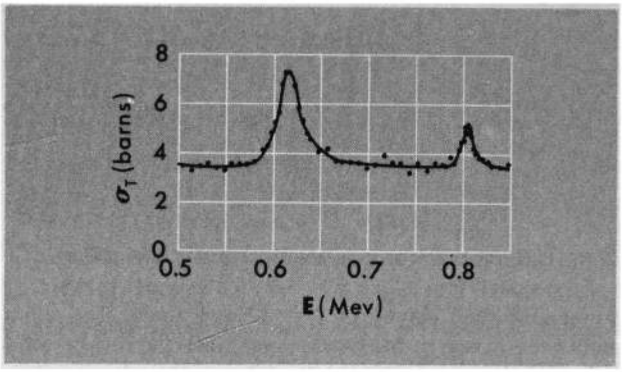
The assumption that a graph must be as smooth as possible can sometimes lead to wrong interpolations, particularly if the graph is constructed from very few experimental points. For example, Fig. 2 – 3 shows what appears to be a reasonable graph of the data represented by the black circles. (The meaning of the symbols used for the quantities being plotted is unimportant for our purposes.) Interpolating from this graph we should predict for $E = 0.62$ Mev that $\sigma_T = 3.5$ barns. You can see how misleading this interpolation is by looking at Fig. 2 – 4, which is a graph of the same experiment but made from many more measurements giving many more experimental points. This graph shows the value of σ_T for $E = 0.62$ Mev to be 7.2 barns, not 3.5! Such functions are relatively uncommon, but that they exist should be borne in mind in making interpolations from a graph.

While one must be careful in interpolating from tables and graphs, extrapolation beyond the ends of a table or graph requires even more caution. This is evident when we try to extrapolate the curve in Fig. 2 – 1(d). The dashed line in Fig. 2 – 5 is a possible extrapolation of the curve in Fig. 2 – 1(d), but, as actual experimental data at large loads show (the dashed rectangles), the dashed line is not in very good agreement with the facts. Extrapolation of some functions beyond an order of magnitude or more can lead to completely erroneous predictions, as you will find out in Section 2 – 8.

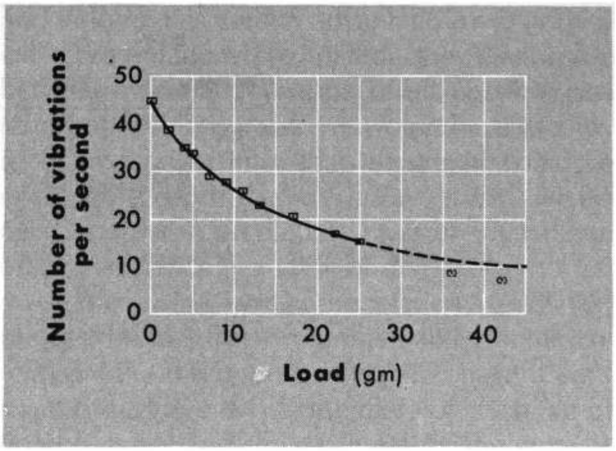
In choosing the proper scales when drawing a graph, one needs to consider the precision of the data. For very precise data, one might want to make a large graph, nearly full-page in size, in order to read to the same precision as that with which the data were determined. With very imprecise data there is no point in drawing a large graph, since only approximate values can be determined from it. There is no need, for ex-



2–3 A graph of the cross section of a beryllium nucleus as a function of the energy of neutrons scattered from the nucleus.



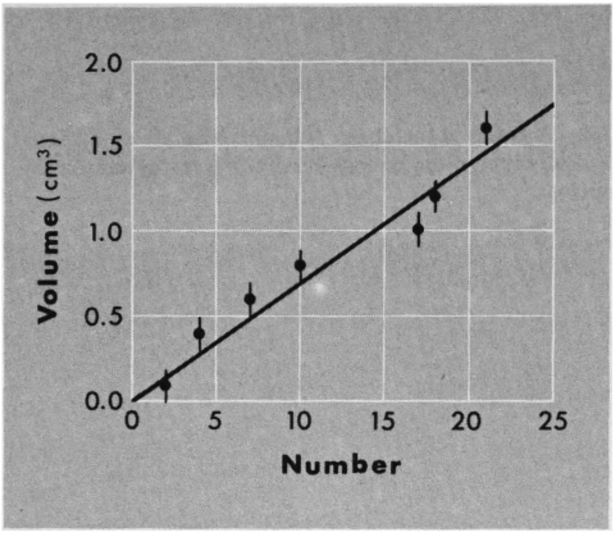
2–4 The same graph as in Fig. 2–3, but drawn from a larger number of experimental points.



2–5 An extrapolation from the graph shown in Fig. 2–1(d). The dashed line represents the extrapolation based on Fig. 2–1(d), and the dashed rectangles are actual experimental data obtained after the extrapolation was made.

ample, to make the graph of the data in Fig. 2 – 6 any larger than is shown in the figure.

To reduce, but not eliminate, the hazards of interpolations and extrapolations, it is best to hunt for a mathematical relation expressed as an equation, which shows how two quantities depend upon each other. If such a relation can be found and checked at a large number of experimental points, then the prediction of values at intermediate points can be made with high confidence and considerable accuracy. In Section 2 – 6 you will see how mathematical relations can often be obtained as generalizations from a graph of experimental data. Many of the laws of physics are expressed most usefully by such mathematical relations, and we shall be constantly seeking such relations throughout this course. In the following sections, we shall discuss several common and useful examples of the ways in which one quantity might depend upon another.



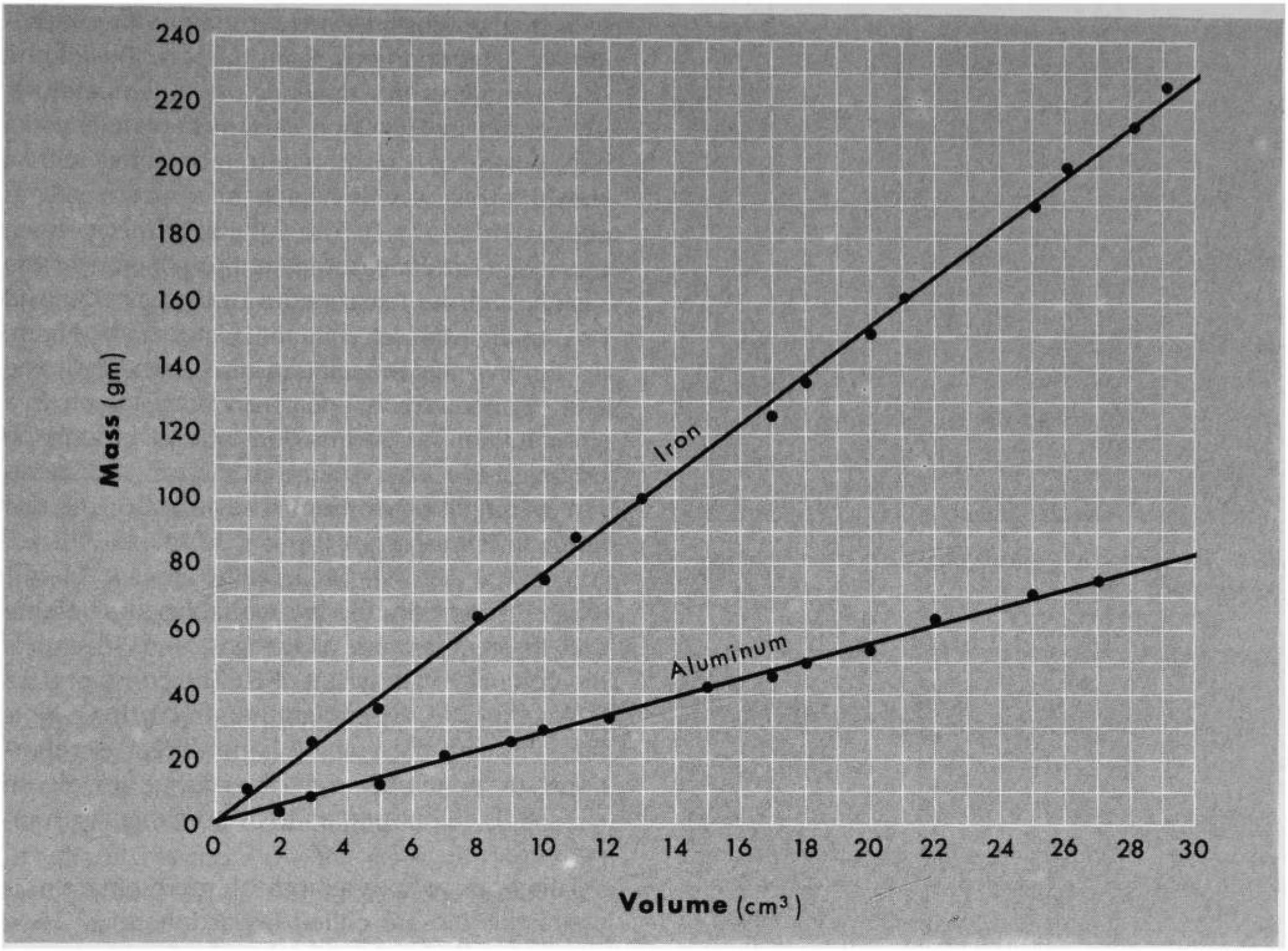
2–6 Different numbers of identical small steel spheres were dropped into a 50 cm³ graduated cylinder of water to measure their volume by water displacement. The graph shows the results of the experiment. Since the number of spheres is known exactly, vertical lines rather than rectangles represent the estimated uncertainty of the measurements.

2–2 DIRECT PROPORTION AND LINEAR FUNCTIONS

One of the simplest functions often encountered in physics is the direct proportion $y = kx$. As an example of a direct proportion consider the relation between the volume and mass of a substance. In a simple experiment, measurements of different-sized pieces of two different metals were made with a simple balance and a centimeter scale, and the results are shown in Table 3. The data have been graphed in Fig. 2 – 7

TABLE 3

IRON		ALUMINUM	
VOLUME	MASS	VOLUME	MASS
(cm^3)	(gm)	(cm^3)	(gm)
1	11	2	4
3	25	3	8
5	36	5	12
8	63	7	21
10	75	9	25
11	88	10	29
13	100	12	33
17	126	15	42
18	136	17	46
20	151	18	50
21	162	20	54
25	189	22	63
26	201	25	71
28	214	27	75
29	226		



2–7 The masses of different-sized samples of iron and aluminum as functions of their volume. The data from which the graphs were plotted is given in Table 3.

and show straight lines through the origin, characteristic of a direct proportion, for both metals. You can see from the graph that the ratio $\frac{M}{V}$ is a constant for any mass of either metal. In other words, $\frac{M}{V} = k$ or $M = kV$, the equation for a direct proportion.

The constant k is called the *proportionality constant*. In our example, as you can see from Fig. 2 – 7, for iron $k_i = 7.7$ gm/cm^3, the density of the iron. We can generalize from this graph and say that for any piece of iron $M = 7.7\ V$ where M is the mass in grams of a sample and V is the volume in cubic centimeters of the sample. Similarly, for aluminum $M = 2.7\ V$. For any homogeneous material whatsoever we can write the expression

$$M = DV$$

where D represents the density of the material in any units consistent with the units used to measure the volume and mass.

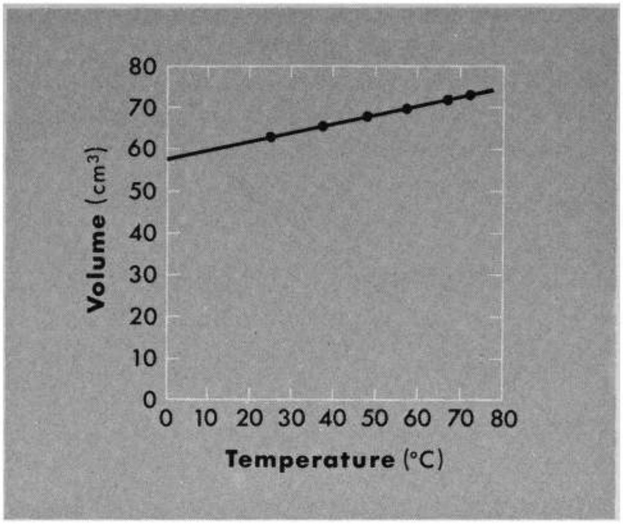
Expressing a function by a general equation such as the one above relating mass, density, and volume has certain advantages over graphical representations. To include the relationship between mass and volume of all possible materials on a single graph, we would have to draw so many straight lines that the graph would be unreadable. With the relation $M = DV$, all we need in order to predict the mass of any volume of any substance is a table of densities.

Another useful form of the relation $M = DV$ expresses the fact that when mass and volume are related by a direct proportion they have a constant ratio. If we divide the mass of a sample of iron by its volume, the result will be the same as that obtained by dividing the mass of any other sample of iron by its volume:

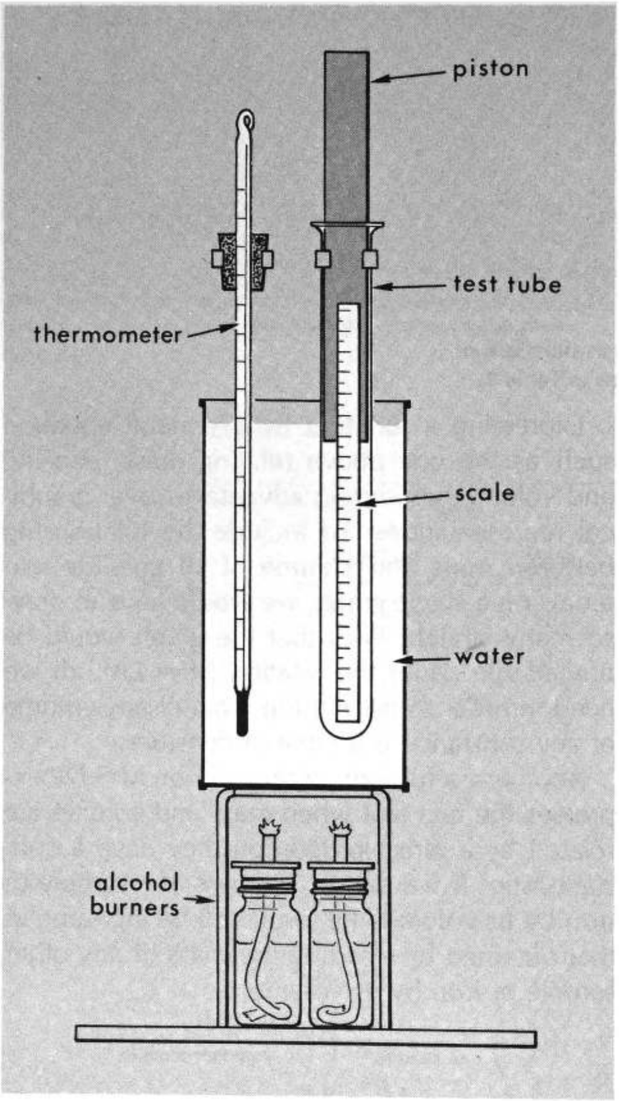
$$\left(\frac{M}{V}\right)_{\text{one sample}} = \left(\frac{M'}{V'}\right)_{\text{another sample}} = D$$

or

$$\frac{M}{V} = \frac{M'}{V'}$$



2–8 The volume of a sample of gas as a function of its temperature.



2–9 Apparatus used to obtain the data for Fig. 2–8.

Not all graphs which yield straight lines represent direct proportions. Consider how the volume of a sample of gas depends on its temperature. The graph in Fig. 2 – 8 shows the results of an experiment in which a test tube fitted with a movable piston was heated in a water bath as shown in Fig. 2 – 9. The volume of the gas as it expanded was measured by the position of the piston, and the temperature of the gas was read on the thermometer in the water bath. Figure 2 – 8 shows that the relationship between volume and temperature is linear, since the graph is a straight line. But the line does not go through the origin; at 0° C the volume is 58 cm^3, and so it is not a direct proportion. The equation of this function is

$$V = 58 + 0.22\,t$$

where V and t are the two variables, the volume and the temperature of the gas, and 58 cm^3 is the volume of the gas at 0° C. The constant 0.22 is the slope of the straight line in cm^3/° C. Note that although the volume is not directly proportional to the temperature, the *change* in volume is directly proportional to the *change* in temperature.

Functions whose graphical representation is a straight line are called linear functions. They can generally be expressed in the form

$$y = y_0 + kx$$

where y_0 and k are constants whose values depend on the physical situation described by the function. Any linear function of this form can be changed into a direct proportion by a suitable selection of the coordinate system that gives $y_0 = 0$.

2–3 POWER LAWS AND SIMILAR FIGURES

Another type of relation occurs when one quantity varies as the square of another. For example, the area A of a square of side L is equal to L^2:

$$A = L^2$$

If L is measured in meters, the area A will be in square meters (m^2). Also, the area A of a circle of radius R is given by

$$A = \pi R^2$$

Both of these equations show that an area varies with the square of a length.

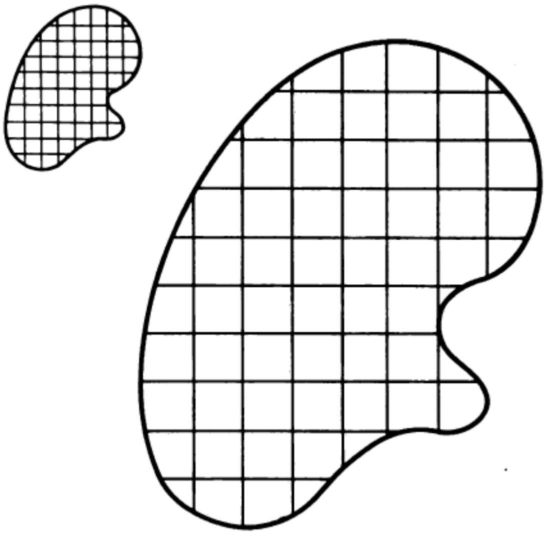
All circles are similar figures: all have the same shape. They are just magnified or reduced copies of one another. Also, all squares are similar figures. But these are not the only similar figures. All sorts of figures may be made in magnified or reduced copies. The two curious areas in Fig. 2 – 10 are similar. One was made by magnifying the other until every linear dimension was tripled. You can check this statement by seeing that the sides of every square in the larger area are just three times what they are in the small one. This means that every square in the large area has nine times the area of the corresponding small one. The total area of the large figure is, therefore, also nine times that of the smaller one. Just as in the case of circles and squares, the areas of any similar figures vary as the square of a linear dimension. When the linear dimension is multiplied by 3, the area is multiplied by 9. In general, then, for similar figures

$$A \propto L^2$$

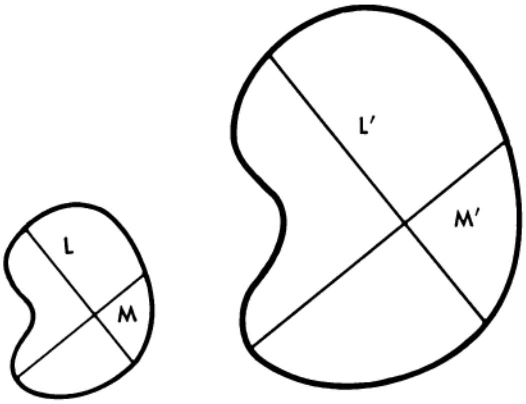
Notice that it does not matter which linear measurement you take for L, as long as you use the corresponding measurement for all the similar figures you are comparing. For example, for a square you could use the diagonal just as well as the side. When the side of one square is n times as long as that of another, the diagonals are in the same ratio. And the area of the first square is n^2 times as large as that of the second. The same thing applies to the corresponding lengths in any two similar figures. (See Fig. 2 – 11.)

Some figures that have the same name may not be similar. Not all rectangles, for example, are similar. We can have two rectangles with the same base b but different heights h. The area is given as the product of the two linear dimensions: $A = bh$. Although such examples are different from similar figures, they have in common the fact that the area is always measured as unit squares, units of the square of a length. If we use the meter as the measure of all lengths, the areas will be specified in terms of the number of meters squared.

Just as all areas are the product of two lengths, all volumes are the product of three linear dimensions. Again we must distinguish between solid figures like cylinders which may have the same base and different heights, and sets of similar solid figures like spheres or cubes in which every linear dimension is magnified or shrunk by the same factor. For similar solid figures, when the linear dimensions are multiplied by the factor n, the volumes are multiplied by the factor n^3, one power of n for each linear



2–10 These two figures are similar; every linear dimension of the larger figure is the same multiple of the corresponding dimension of the smaller. In this case the dimensions of the larger figure are three times those of the smaller. You can check this by measuring corresponding squares.



2–11 Because these two figures are similar, the ratio of their areas is the same as the ratio of the squares of any two corresponding dimensions. Here, since L' is twice L (M' is also twice M), the area of the larger figure is four times the area of the smaller.

dimension. For example, the volume of a sphere is

$$V = \frac{4\pi}{3} R^3$$

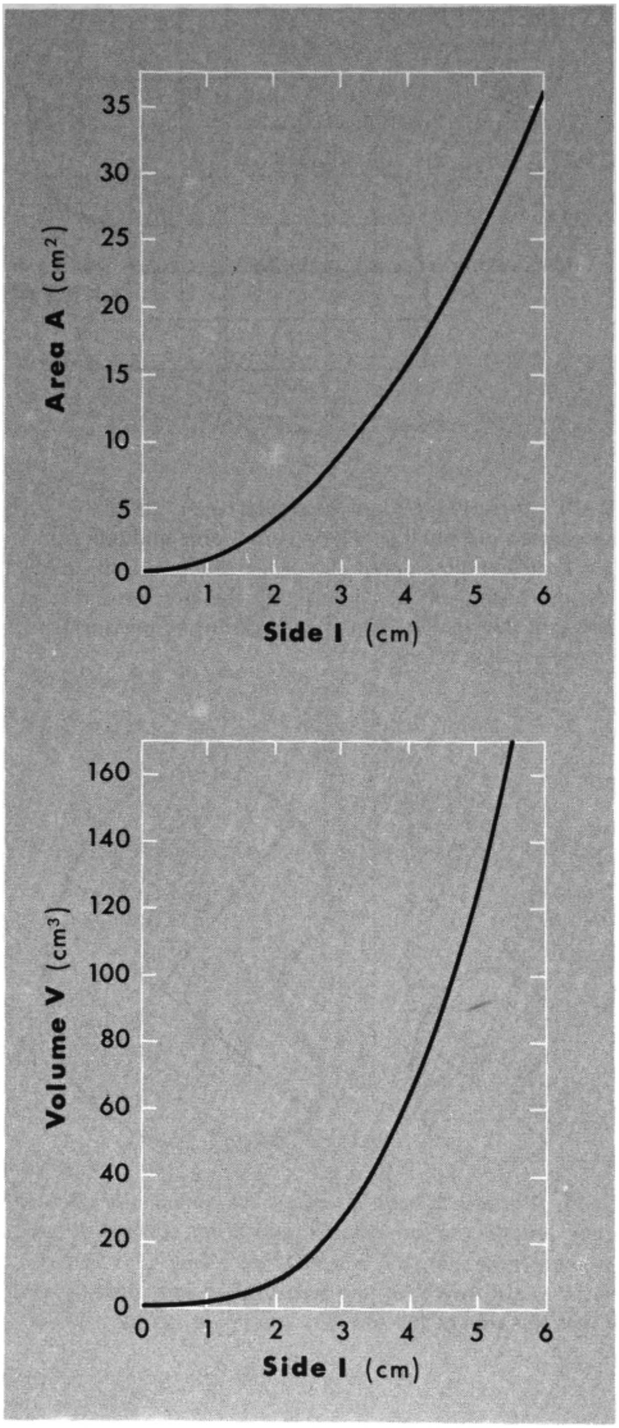
where R is its radius. Then a sphere of radius $R' = nR$ has a volume

$$\begin{aligned} V' &= \frac{4\pi}{3}(R')^3 \\ &= \frac{4\pi}{3} n^3R^3 \\ &= n^3V \end{aligned}$$

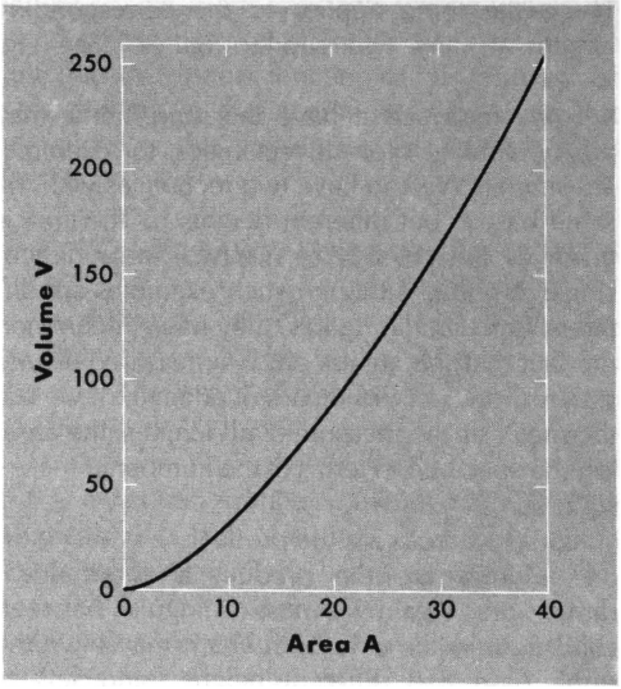
This is just a particular example of the general rule: the ratio of the volumes of similar solids is the cube of the ratio of their linear dimensions.

Relations in which one quantity is proportional to a power, such as the square, cube, etc., of another occur often in physics. They are called power laws. In addition to first, second, and third power laws, such as $M = DV$, $A = kL^2$, $V = kL^3$, which have just been discussed, we also find inverse-power laws, such as $I = \frac{k}{d^2}$ —which means that the intensity of light is inversely proportional to the second power of the distance from the light source. We shall discuss the inverse-square relation in the next section.

Expressing functions as algebraic relations or equations is especially useful if we wish to combine two equations. For example, we might ask the question, how is the area of one face of a cube related to its volume? Figure 2 – 12 shows graphically how each is related to the length of the side, but it would be a laborious process to combine these two graphs to find how the area is related to the volume. Knowing the equations



2–12 Graphs of the area of one face of a cube and of the volume of a cube as functions of the length of one side.



2–13 The volume of a cube as a function of the area of one face.

for area and volume of a cube in terms of its side makes the job easy. Expressing the length of the side in terms of the area and volume respectively, we have

$$l = \sqrt{A}$$
$$l = \sqrt[3]{V}$$

therefore,

$$\sqrt[3]{V} = \sqrt{A}$$
$$V = (\sqrt{A})^3 = A^{3/2}$$

The graph of this function is shown in Fig. 2 – 13.

2–4 THE INVERSE-SQUARE RELATION AND STELLAR DISTANCES

If the light from a small source spreads out equally in all directions (which is very nearly true for a street lamp and a star and many other sources), it can be pictured as shown in Fig. 2 – 14. Here we consider just a portion of the light moving out through a sort of "pyramid" from the point *P*. As the distance from the source increases, the light spreads over a greater area. We define the amount of light passing per unit time through a unit area perpendicular to the direction in which the light is moving as the intensity of the light.

The side of each of the squares in Fig. 2 – 14 is proportional to its distance from *P*. Therefore, the area of each square is proportional to the square of that distance. If we call the distance *d*, and the area the light passes through *A*, this can be expressed as

$$A \propto d^2$$

Hence *I*, the intensity of the light at a distance *d* from the source, is inversely proportional to d^2,

$$I \propto \frac{1}{d^2}$$

This inverse-square relation can also be expressed by the equation

$$\frac{I'}{I} = \frac{(d)^2}{(d')^2} = \left(\frac{d}{d'}\right)^2$$

Note that this relation holds for a single source at distances d^1 and d. It also holds for two identical sources at different distances. For example, suppose we have two identical street lamps, at different distances d_1 and d_2 away from us. Then, the intensities of the light from the two are in the ratio

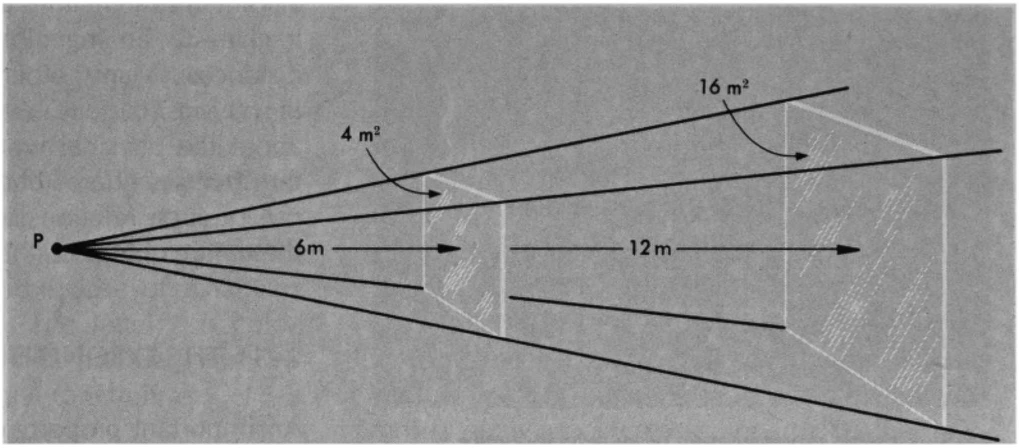
$$\frac{I_1}{I_2} = \frac{{d_2}^2}{{d_1}^2}$$

This relation enables us to estimate the distance of one lamp if we have another identical lamp at a known distance. For example, suppose we find that a lamp 10 meters (d_1) away gives an intensity that is 16 times that of an identical lamp at some unknown distance d_2. (Photoelectric cells, camera light meters, and photographic plates can give accurate measures of relative intensity.)

$$16 = \frac{I_1}{I_2} = \frac{{d_2}^2}{(10 \text{ meters})^2}$$

$$d_2 = \sqrt{16 \times (10 \text{ meters})^2} = 40 \text{ meters}.$$

We can make a graph of the intensity of the



2–14 The inverse-square relation. Light from a point (P) radiates in all directions. Since the light spreads out to cover four times the area at twice the distance, it follows that it can be only $\frac{1}{4}$ as intense. Thus, when the distance is doubled, the intensity decreases to $\frac{1}{4}$, or the intensity is inversely proportional to the square of the distance.

light from a lamp versus the distance from the lamp. Let us say the intensity at 10 meters is I_1. As we have just found, the intensity at 40 meters is $I_1/16$. We could make the following table by making a few more similar calculations:

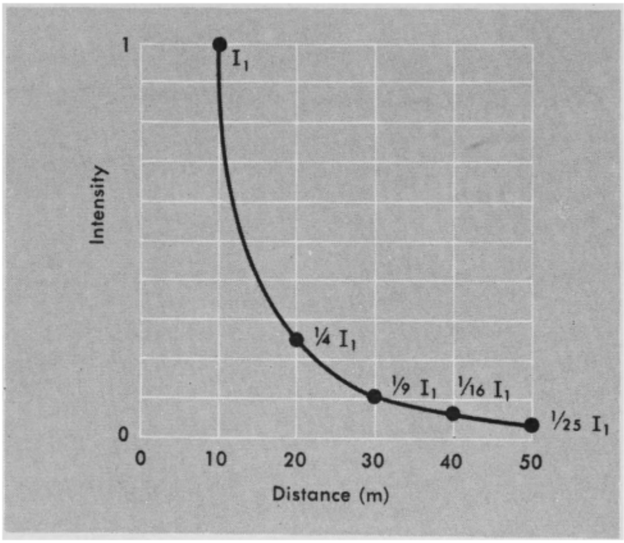
Intensity of Light as a Function of Distance from Lamp

DISTANCE (m)	INTENSITY
10	I_1
20	$I_1/4$
30	$I_1/9$
40	$I_1/16$
50	$I_1/25$

The graph of this table is shown in Fig. 2 – 15.

The method used to find the relative distances of the two street lamps can be used to find the distances of stars so far away that their distances cannot be measured by triangulation using the diameter of the earth's orbit as a base line. The measurement is made by comparing the intensity of the faint image of a far-away star on a photographic plate with the intensity of a nearby star which appears to give off the same amount of light.

We can see how the inverse-square relation works by using it to measure the distance of a nearby star, α Centauri A, whose distance we can measure geometrically by triangulation. Judging from its color and calculated mass, this star is very similar to the sun. But the intensity of illumination here at the earth is 10^{11} times greater from the sun than from α Centauri A. From the inverse-square relation we learn that α Centauri A must be about $\sqrt{10^{11}} = 3 \times 10^5$ times as far away from us as the sun is. The sun is 1.5×10^{11}m away, so the star must be about 4.5×10^{16}m away. And this is almost exactly what a geometrical measurement also tells us. This agreement of the inverse-square method for measuring stellar distances and the triangulation method for stars that are nearby strengthens our confidence when we apply the inverse-square relation to the measurement of distances too great to be triangulated. The measurement is at best a rough one, because we do not expect that the two stars are really equally strong sources of light. But in this rough way we can go far beyond the possibilities of triangulation methods and at least determine the order of magnitude of the distance to far-off stars.



2–15 Graphic representation of an inverse-square relation – the intensity of light versus the distance from the lamp.

The inverse-square method of distance measurement has other limitations. The inverse-square relation will certainly not apply if there is anything between the eye and the source that makes light deviate from a straightline path or absorbs some of it. Obviously, a patch of fog would reduce the intensity from a distant lamp and upset any calculations based on the inverse-square law.

The inverse-square relation describes many situations in nature where something – light, particles, or lines of electric force – radiates from a point in straight lines uniformly in all directions. It gives us an ingenious way to measure great distances. Many other mathematical relations in physics can be exploited to tell us things about the physical world. We have introduced the inverse-square relation here to illustrate the use of such relations in physics, not to discuss the nature of light. Light will be our topic later on.

2–5 THE EXPONENTIAL FUNCTION

An important property of living cells is that they can reproduce their own kind and thus increase in number with time. The number of new cells produced in a given interval of time depends on the number of cells present at the beginning of the time interval. In other words, the number of new cells produced in a given time interval is proportional to the number present. Consider, for example, a colony of bacteria that increases in number by 10 percent every hour. If, at a certain time, there are 100 of them, at the end of one hour there will be 110. After two hours there will be $110 + 0.1\ (110) = 121$, and so on. After twelve hours there will be 315, and after twenty-four hours there will be 985. Figure 2 – 16 is a graph of the size of the bacteria colony as a function of time. Notice that the rate of growth of the colony increases rapidly with time and the size doubles every 7.3 hours. Of course, a doubling in the size of the colony every 7.3 hours could not continue for very long. Soon the available food supply would limit the increase, and the population would stop growing.

There are many physical changes in which the change in a quantity in a given time interval is proportional to the magnitude of the quantity at the start of the interval. Figure 2 – 17, for example, is a graph of the results of an experiment in throwing dice. Thirty pairs of dice were shaken together in a box and then spilled out on a table. All the dice that showed a five were picked up and separated from the rest. The remaining dice were shaken again in the box and again spilled out on the table and again the fives were removed. Twelve such throws were made, and Fig. 2 – 17 shows the number of dice remaining on each throw after the fives were removed.

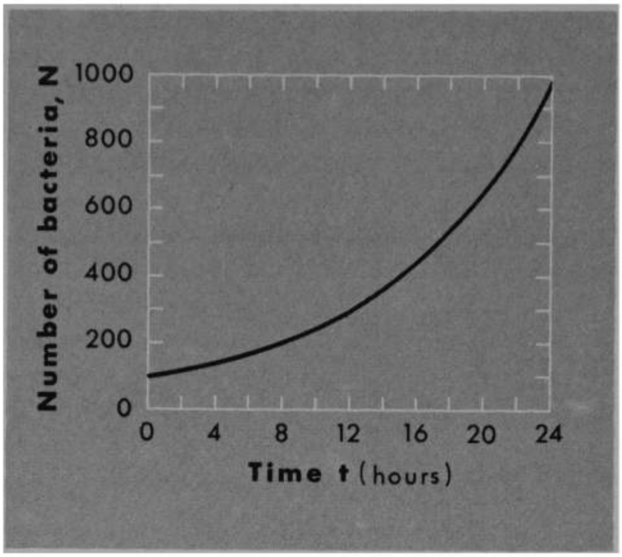
Functions like those illustrated in Figs. 2 – 16 and 2 – 17 are called exponential functions and have the form

$$y = y_0 a^{kx}$$

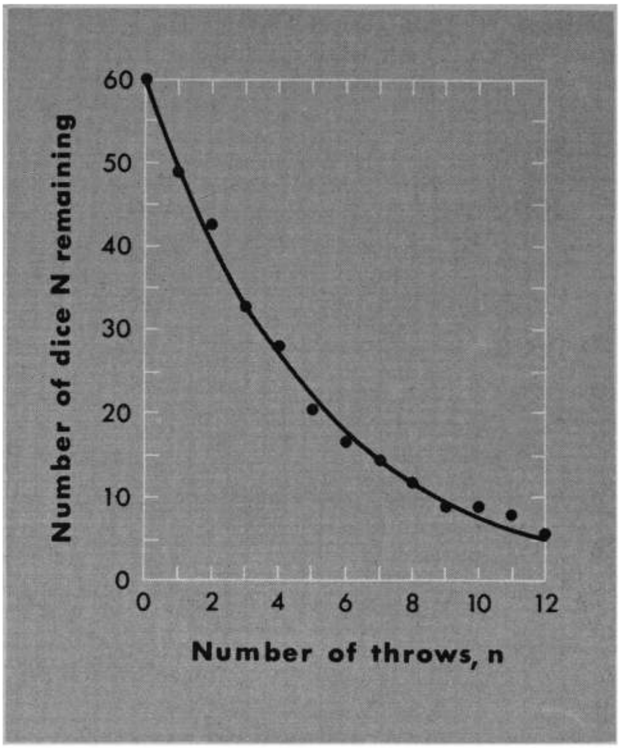
in which y_0 is the initial value of y, k is a number determined by the rate of change, and a is a constant number. Actually, a can be any positive number other than 1, but for the purposes of mathematical analysis it is most convenient to choose $a = 2.718\ldots$, the base of what are called natural logarithms. The letter e is commonly used to represent this number, and we write the simple exponential function as

$$y = y_0 e^{kx}$$

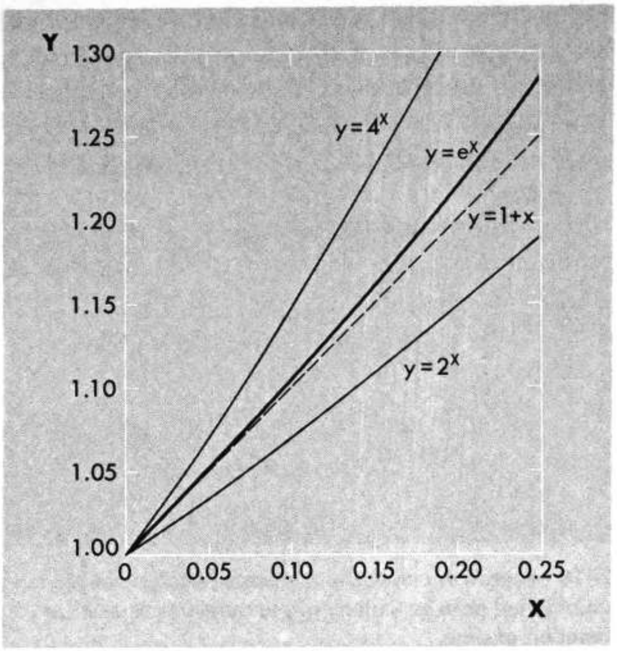
Figure 2 – 18 shows one reason why 2.718 is a useful base for the exponential function. Here the function $y = a^x$ has been plotted for different values of a. Although the lines begin to slope up more steeply as x increases, for small values of x, all three lines are very nearly straight, and the line for $a = e$ has a very useful property. As you



2–16 A growth curve: the number of bacteria in an uninhibited growing colony of microorganisms as a function of time.



2–17 A curve resulting from a simple experiment with dice.



2–18 A graph of $y = a^x$ for small values of x and different values of a. For $y = e^x$, where $e = 2.718 \ldots$, the graph very nearly coincides with the line representing $y = 1 + x$, when x is small.

can see from the graph, for very small values of x, $y = e^x \approx 1 + x$. Taking the natural logarithm of both sides of this equation, we have $\ln e^x = x \ln e \approx \ln(1 + x)$. (We write "ln" for the natural logarithm to distinguish it from "log," which we use as the symbol for logarithms to the base 10.) Since e, by definition, is the base of the natural logarithms, $\ln e = 1.000$ and we have

$$\ln(1 + x) \approx x$$

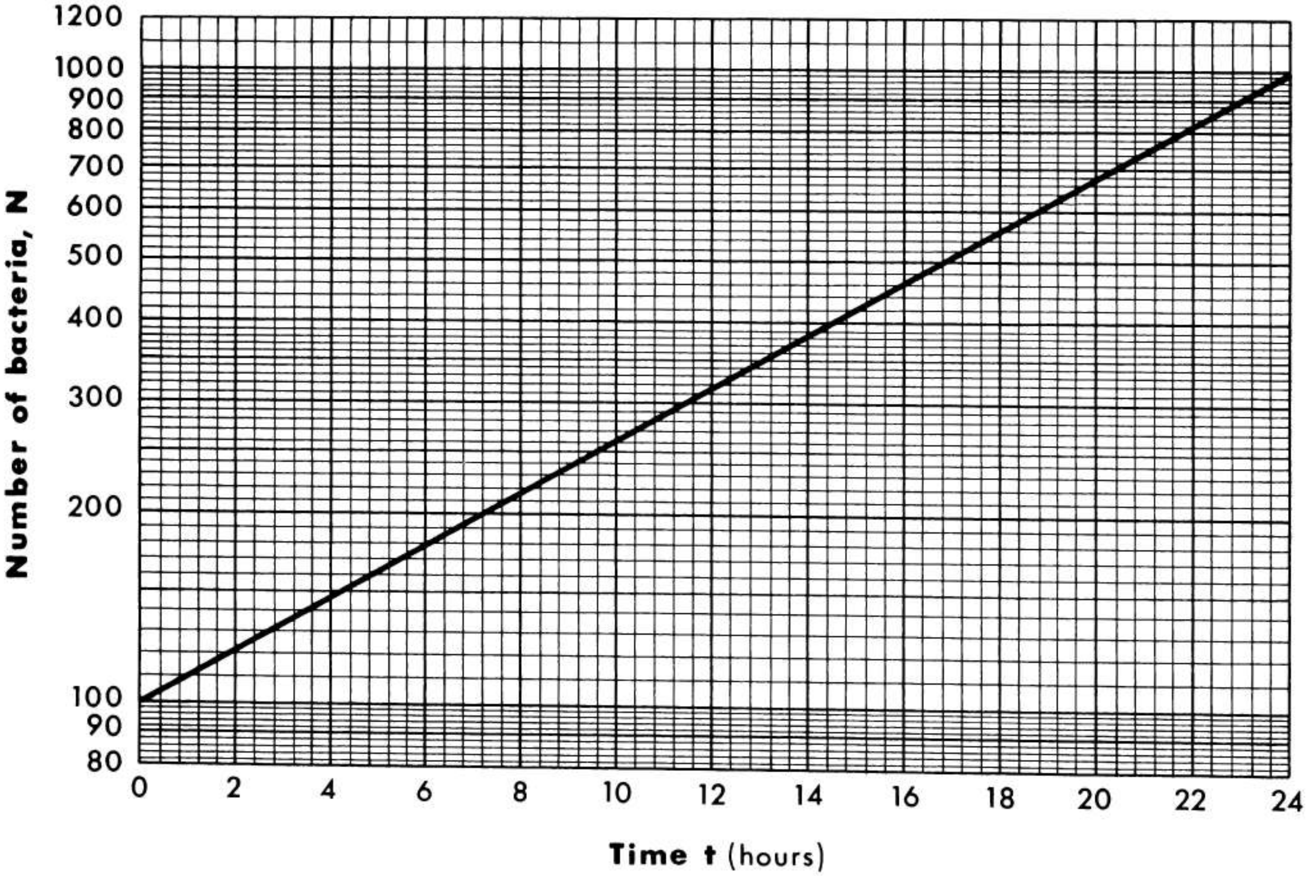
This approximation for the natural logarithm of a number very close to 1 will be useful later.

2–6 THE EXPONENTIAL FUNCTION AND SEMI-LOG PLOTS

It is sometimes convenient to write the exponential function in logarithmic form. Taking the natural logarithm of both sides of $y = y_0 e^{kx}$ gives

$$\ln y = \ln y_0 + kx$$

and this equation gives a straight line when ln y

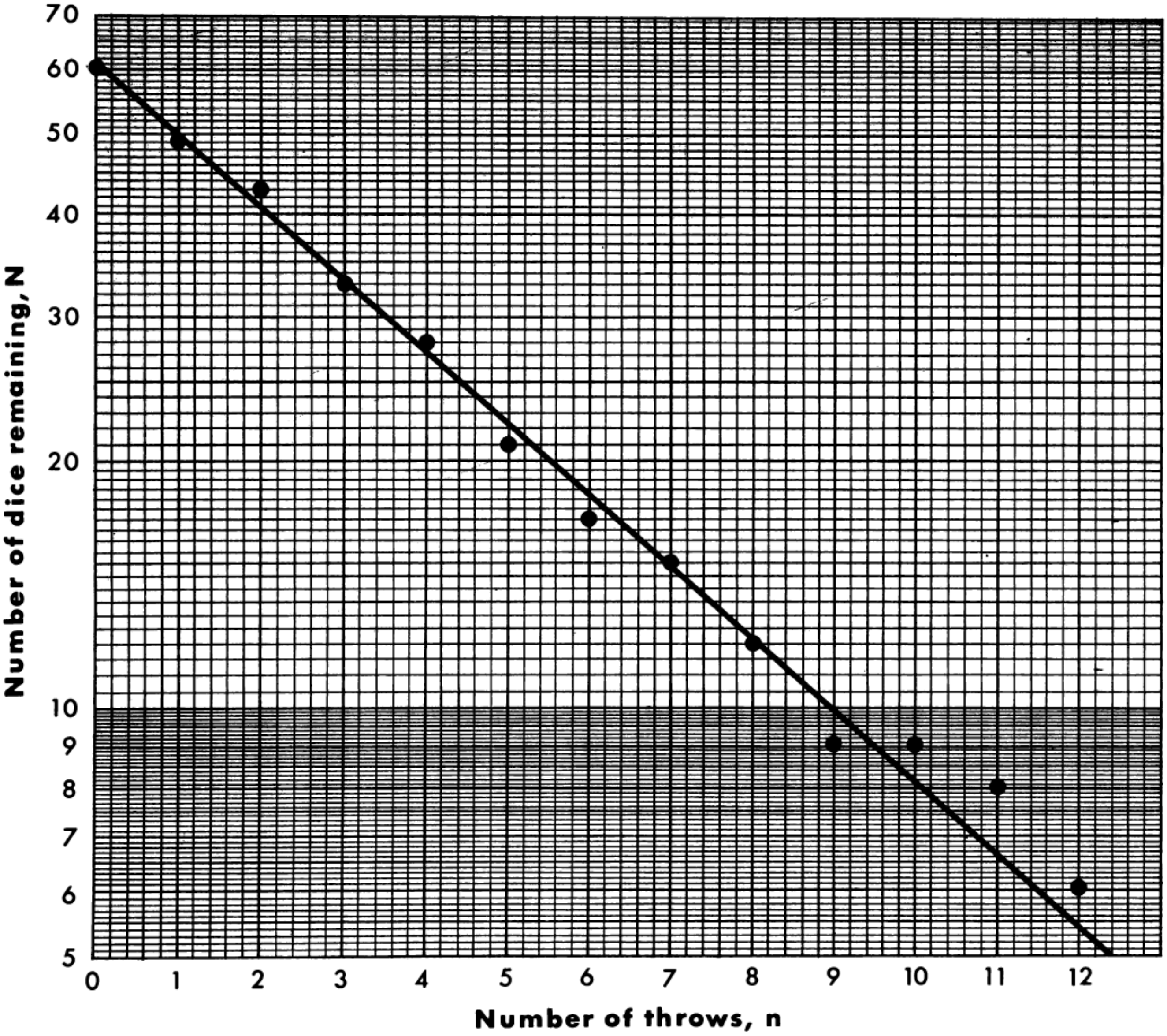


2–19 A semi-log plot of the number of bacteria in an uninhibited growing colony of microorganisms as a function of time gives a straight line.

is plotted as a function of x. This will also be true if we use logarithms to the base 10 instead of the natural logarithms, because the logarithm of any number to one base is proportional to the logarithm to another base. Therefore, if a plot of ln y as a function of x yields a straight line, so will a plot of log y as a function of x, but with a different slope. This fact is very useful when we try to find an analytical expression for a graph which was obtained from experimental data. If we suspect an exponential relationship, we try a plot of the logarithm of one variable as a function of the other variable. If the guess is right, we get a straight line.

Fortunately, we can make such graphs, which are called *semi-log* plots, without having to look up all the logarithms of one variable in a table of logarithms. We can use graph paper which is scaled logarithmically along one axis (like the "C" and "D" scales on a slide rule) and linearly along the other axis (like a ruler). Figures 2 – 19 and 2 – 20 are semi-log plots of Figs. 2 – 16 and 2 – 17 respectively.

Semi-log plots have another application in addition to searching for analytical expressions. Often we are interested in a graphical representation of a function where one variable changes over several orders of magnitude, while the other



2–20 A semi-log plot of the number of dice remaining in a simple experiment with dice as a function of the number of throws.

does not. Consider the relation between the height above sea level and the density of the atmosphere. An ordinary graph made from a table of data yields Fig. 2 – 21. While this figure will do for reading off values below 20 km, it is completely useless above 40 km. Figure 2 – 22 is a graph of the same data with the density plotted on a logarithmic scale. It can clearly be used over the entire range.

2–7 POWER LAWS AND LOG-LOG PLOTS

In Sections 2 – 2 to 2 – 4 we discussed several different power laws; relations in which some quantity varies, directly or inversely, as some power of another quantity. Such power laws and many others can be represented by the general equation

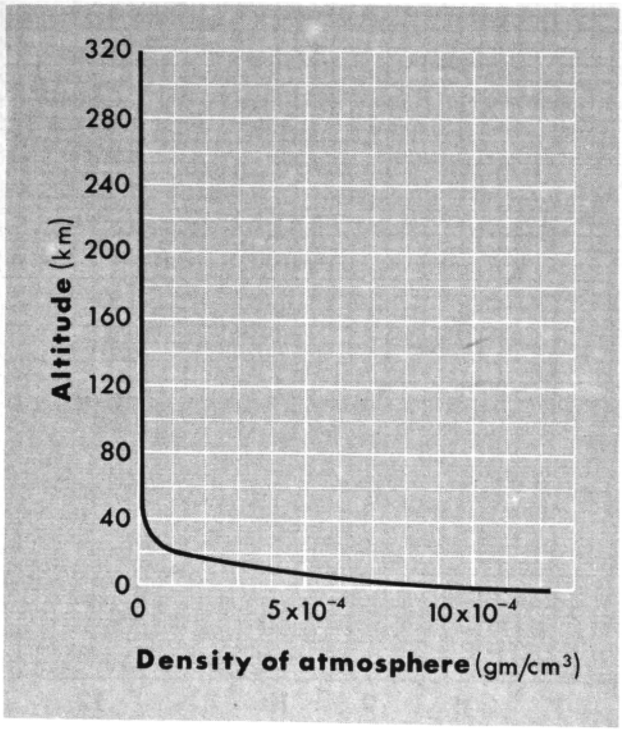
$$y = kx^n$$

where the power, n, is any number, positive or negative. Suppose we do an experiment and get a table of data relating one quantity to another and suspect the relation between the quantities is a power law. It is difficult to tell from a direct graph which is a curved line whether or not we have a power law, and, if so, what is the actual power that we are dealing with—what n is in the relation $y = kx^n$. There is an easy way to find out by the use of logarithms. If we take the logarithm of both sides of the general power law, $y = kx^n$, we get

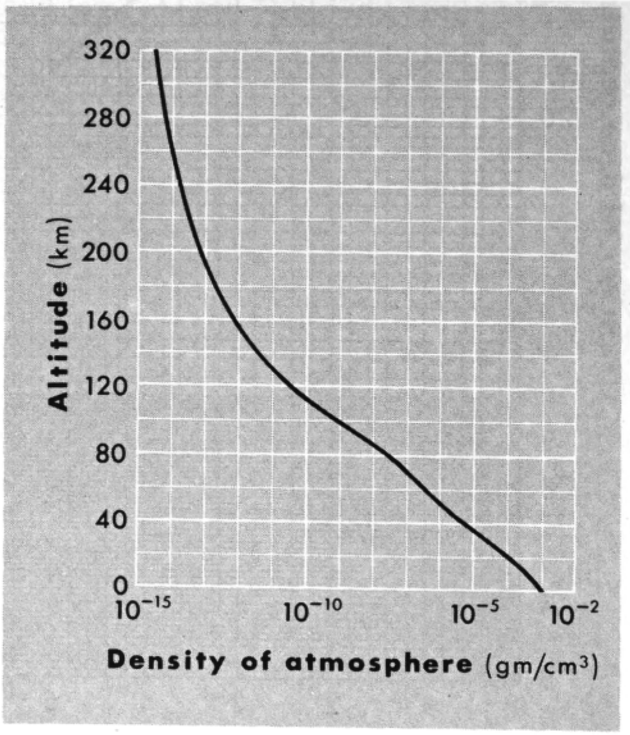
$$\log y = \log k + n \log x$$

Since this is a linear relationship between log y and log x, it will give a straight line when log y is graphed as a function of log x, with the slope determined by n. Just as in the case of the semi-log plots, we do not have to look up the logarithm of either variable. Instead we use graph paper that is scaled logarithmically along both axes (log-log paper).

Suppose, for example, we make a graph from data relating the time for one complete swing of a pendulum to the length of the pendulum. The result is shown in Fig. 2 – 23. It is not easy to determine from this graph if the relationship is a power law or not, let alone the order of the power (the value of n in $y = kx^n$) or the value of k. However, if we plot T as a function of L on log-log paper as shown in Fig. 2 – 24, we see that we get



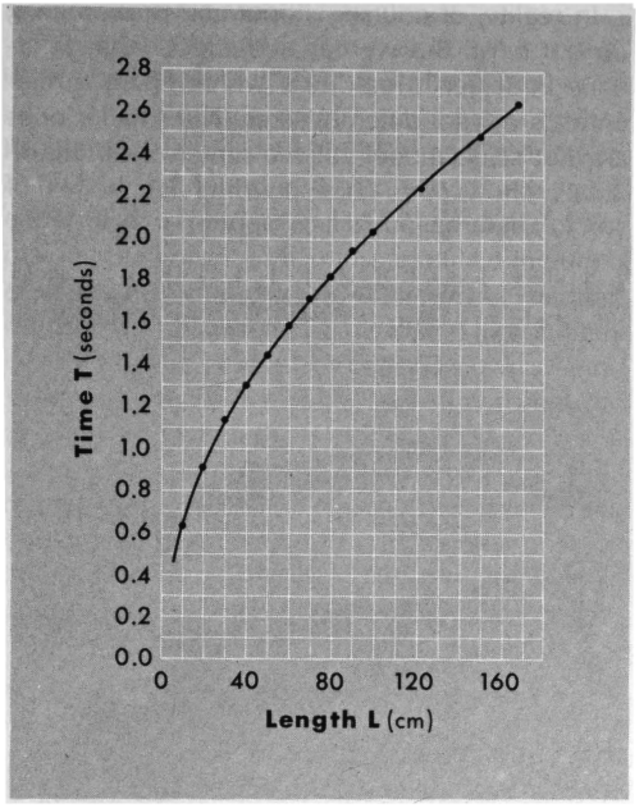
2–21 The density of the earth's atmosphere plotted as a function of altitude above the earth's surface. The graph is of little use for altitudes above about 40 km.



2–22 When the data for atmospheric density used to plot the graph in Fig. 2–21 is used to make a semi-log plot, the graph is useful to altitudes as great as 320 km.

a straight line, showing a linear relation between log L and log T, and this tells us that L and T are related by a power law. Furthermore, if we measure a and b in Fig. 2 – 24 in centimeters, we find the slope of the graph, a/b, is 1/2. This is the value of n in the equation $\log T = \log k + n \log L$ for the straight line. Hence, $T = kL^{1/2} = k\sqrt{L}$. To find the proportionality constant k it is enough to read off one pair of points on the graph, L_1 and T_1, then

$$k = \frac{T_1}{\sqrt{L_1}}$$

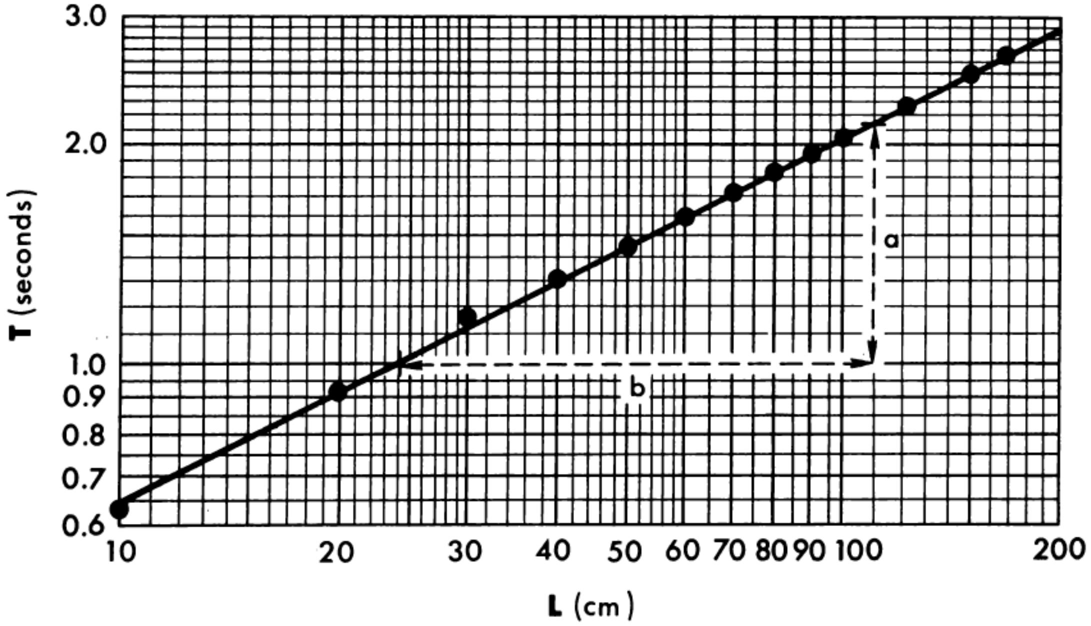


2–23 The time for one complete swing of a simple pendulum as a function of its length.

The power functions and the exponential function are very common in physics, but there are many other functions as well which do not yield a straight line when plotted logarithmically. But you will find that logarithmic graphs will help you a great deal in analyzing much of the experimental data you will obtain in the laboratory.

2–8 SCALING—THE PHYSICS OF LILLIPUT

The fictional traveler Lemuel Gulliver spent a busy time in a kingdom called Lilliput, where all living things—men, cattle, trees, grass—were exactly similar to the corresponding things in our world, except that they were all built on the scale of about 1 to 10. Lilliputians were a little under 20 cm high, on the average, and built proportionately just as we are. Gulliver also visited Brobdingnag, the country of the giants, who were exactly like men but about ten times as tall. As Swift described it, daily life in both kingdoms was about like ours (in the eighteenth century). His commentary on human behavior is still worth reading, but we shall see that people



2–24 A log-log plot of the function graphed in Fig. 2–23.

of such sizes just could not have been as he described them.

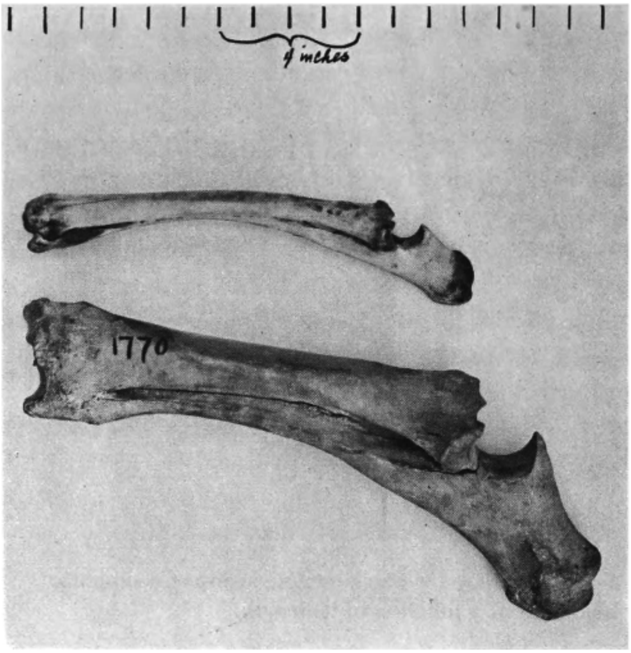
Long before Swift lived, Galileo understood why very small or very large models of man could not be like us, but apparently Dean Swift had never read what Galileo wrote. One character in Galileo's *Two New Sciences* says, "Now since . . . in geometry, . . . mere size cuts no figure, I do not see that the properties of circles, triangles, cylinders, cones, and other solid figures will change with their size. . ." But his physicist friend replies, "The common opinion is here absolutely wrong." Why?

We start with the strength of a rope. It is easy to see that if one man who pulls with a certain strength can almost break a certain rope, two such ropes will just withstand the pull of two men. A single large rope with the same total area of cross section as the two smaller ropes combined will contain just double the number of fibers of one of the small ropes, and it will also do the job. In other words, the breaking strength of a wire or rope is proportional to its area of cross section, or to the square of its diameter. Experience and theory agree in this conclusion. Furthermore, the same relation holds, not only for ropes or cables supporting a pull, but also for columns or struts supporting a thrust. The thrust which a column will support, comparing only those of a given material, is also proportional to the cross-sectional area of the column.

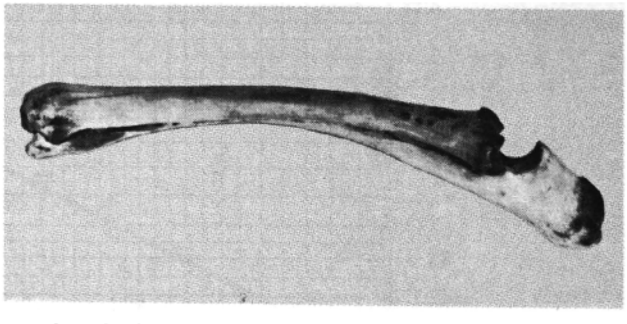
Now the body of a man or an animal is held up by a set of columns or struts—the skeleton—supported by various braces and cables, which are muscles and tendons. But the weight of the body which must be supported is proportional to the amount of flesh and bone present, that is, to the volume.

Let us now compare Gulliver with a Brobdingnagian giant, ten times his height. Since the giant is exactly like Gulliver in construction, every one of his linear dimensions is ten times the corresponding one of Gulliver's. Hence, the bodies of the giant and of Gulliver are exactly similar in cross section and in shape. Because the strength of the giant's columns and braces is proportional to their cross-sectional area which is $(10)^2$ times the cross-sectional area of Gulliver's bones, his bones will be 100 times as strong as Gulliver's. Because his weight is proportional to his volume, which is $(10)^3$ times Gulliver's volume, it will be 1000 times as great as Gulliver's. So the giant will have a strength-to-weight ratio only 1/10 of ours. Just to support his own weight, he would have as much trouble as we should have in carrying nine men on our back.

In reality, of course, Lilliput and Brobdingnag do not exist. But we can see effects of a difference in scale in real life if we compare similar animals of very different sizes. The smaller ones are not scale models of the larger ones. Figure 2–25 shows the corresponding leg bones of



2–25a The front leg bones of a bison and a gazelle. The animals are related, but the gazelle is much smaller. The photos show the approximate relative sizes of the bones.



2–25b The leg bone of a gazelle enlarged to the same length as the bison bone. Note that the bone of the larger animal is much thicker in comparison to its length than that of the gazelle. The small deer is generally more lightly and gracefully built. Can you visualize how much different Lilliputians must have been from men of normal size?

two closely related animals of the deer family: one a tiny gazelle, the other a bison. Notice that the bone of the large animal is not at all similar geometrically to that of the smaller. It is much thicker for its length, thus counteracting the scale change, which would make a strictly similar bone too weak.

Galileo wrote very clearly on this very point, disproving the possibility of Brobdingnag, or of any normal-looking giants: ". . . if one wishes to maintain in a great giant the same proportion of limb as that found in an ordinary man he must either use a harder and stronger material for making the bones, or he must admit a diminution of strength in comparison with men of medium stature; for if his height be increased inordinately he will fall and be crushed under his own weight. Whereas, if the size of a body be diminished, the strength of that body is not diminished in the same proportion; indeed, the smaller the body the greater its relative strength. Thus a small dog could probably carry on his back two or three dogs of his own size; but I believe that a horse could not carry even one of his own size." Figure 2 – 26 is taken from a drawing by Galileo, who drew it to illustrate the paragraph just quoted.

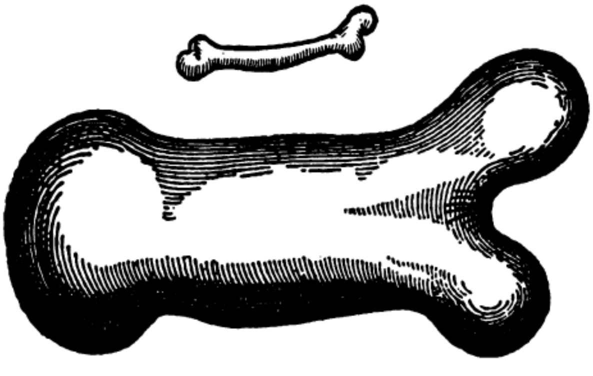
An elephant is already so large that his limbs are clumsily thickened. However, a whale, the largest of all animals, may weigh forty times as much as an elephant; yet the whale's bones are not proportionately thickened. They are strong enough because the whale is supported by water. What is the fate of a stranded whale? His ribs break. Some of the dinosaurs of old were animals of whalelike size; how did they get along?

Following Galileo, we have investigated the problems of scaling up to giants. Now let's take a look at some of the problems that arise when we scale down.

When you climb dripping wet out of a pool there is a thin film of water on your skin. Your fingers are no less wet than your forearm; the thickness of the water film is much the same over most of your body. Roughly, at least, the amount of water you bring out is proportional to the surface area of your body. Perhaps you carry out of the pool a glassful or so, which amounts to about a 1 percent increase in what you have to move about. A Lilliputian who is 1/10 as tall as you and exactly similar will have a weight that is only $(1/10)^3$ of yours. However, his surface area will be $(1/10)^2$ of yours, so the amount of water he carries out of a pool will be $(1/10)^2$ of the amount you carry out. Hence, for him, the ratio (amount of water)/(original weight) is 10 times as great as for you. He carries out about 10 percent of his weight, which would be equivalent to a heavy winter suit of clothing and an overcoat.

There is a still more important effect of the scale of a living body. Your body loses heat mainly through the skin (and some through breathing out warm air). It is easy to believe—and it can be checked by experiment—that the heat loss is proportional to the surface area, keeping other factors, like the temperature and the nature of skin, constant. The food taken in must supply this heat, as well as a surplus amount of energy we use in moving about. So minimum food needs are proportional to surface area. If a man like Gulliver can live off a leg of lamb and a loaf of bread for a day or two, a Lilliputian with the same body temperature will require only $(1/10)^2$ this volume of food. But his leg of lamb, scaled down to his world, will be smaller in volume by a factor of $(1/10)^3$. Therefore, he would need ten of his roasts and loaves to feel as well fed as Gulliver did after one. Lilliputians would have to be a hungry, restless lot. You can recognize these qualities in many small mammals, like a mouse.

We can see why there are no warm-blooded animals much smaller than the mouse. Fish and frogs and insects can be very much smaller because their body temperature is not higher than that of their surroundings. In accord with the



2–26 Galileo's drawing illustrating scaling. Over 300 years ago, Galileo wrote concerning the fact that a bone of greater length must be increased in thickness in greater proportion in order to be comparatively strong.

scaling laws of area and volume, small warm-blooded animals need, relatively, a great deal of food; very small ones could not gather or even digest such an enormous amount. Certainly the agriculture of the Lilliputians could not have supported a kingdom like the one Gulliver described.

Now we see that neither Brobdingnag nor Lilliput can really be a scale model of our world. But what have these conclusions to do with physics?

Let's start again with the very large. As we scale up any system, the load will eventually be greater than the strength of the structure. This effect applies to every physical system, not just to animals, of course. Buildings can be very large because their materials are stronger than bone, their shapes are different, and they do not move. These facts determine the constants like k in the equation

$$\text{strength} = k\ (\text{thickness})^2$$

but the same laws hold. No building can be made which will look like the Empire State Building in New York and be as high as a mountain, say ten thousand meters. Mountains are solid structures, for the most part, without interior cavities. Just as the bones of a giant must be thick, an object of mountainous size on the earth must be all but solid, or else built of new materials yet unknown.

Our arguments are not restricted to the surface of the earth. We can imagine building a tremendous structure far out in space away from the gravitational pull of the earth. The load then is not given by the earth's gravitational pull, but —as we shall see later—as the structure is built larger and larger each part pulls gravitationally on every other part and soon the outside of the structure is pulled in with great force. The inside, built of ordinary materials, is crushed, and large protuberances on the surface break off or sink in. As a result, any large structure like a planet has a simple shape, and, if it is large enough, the shape is close to that of a sphere. Any other shape will be unable to support itself. Here, then, is the essential reason why the planets and the sun tend to be spherical. The pull of gravity is important for us on earth, but as we extend the range of dimensions which we study, it becomes absolutely dominant in the very large. Only motion can change this result. The great masses of gas which are nebulae, for example, are changing in time, and hence the law that large objects must be simple in shape is modified.

When we go from our size to the very small, gravitational effects cease to be important. But, as we saw in investigating Lilliput, surface effects become significant. If we go far enough toward the very small, surfaces no longer appear smooth, but are so rough that we have difficulty in defining a surface. Other descriptions must be used. In any case, it will not come as a complete surprise that in the domain of the atom, the very small, scale factors demonstrate that the dominant pull is one which is not easily observed in everyday experience.

Such arguments as these run through all of physics. Like order-of-magnitude measurements, they are extremely valuable when we begin the study of any physical system. How the behavior of a system will change with changes in the scale of its dimensions, its motion, and so on, is often the best guide to a detailed analysis.

Even more, it is by the study of systems built on many unusual scales that physicists have been able to uncover unsuspected physical relations. When changing scale, one aspect of the physical world may be much emphasized and another one may be minimized. In this way we may discover, or at least get a clearer view of, things which are less obvious on our normal scale of experience. It is largely for this reason that physicists examine, in and out of their laboratories, the very large and the very small, the slow and the rapid, the hot and the cold, and all the other unusual circumstances they can contrive. In examining what happens in these circumstances we use instruments both to produce the unusual circumstances and to extend our senses in making measurements.

It is hard to resist pointing out how much the scale of man's own size affects the way he sees the world. It has been largely the task of physics to try to form a picture of the world which does not depend upon the way we happen to be built. But it is hard to get rid of these effects of our own scale. We can build big roads and bridges which are long and thin, but are essentially not three-dimensional, complex structures. The very biggest things we can make which have some roundness, which are fully three-dimensional, are buildings and great ships. These lack a good deal of being a thousand times larger than men in their linear dimensions.

Nor can we, as yet, build anything complex like a watch or a radio which is scaled down to 10^{-4} of our own length. In this range of magnitudes lies all our engineering. Physics goes much beyond, outward to the galaxies and inward to the nucleus of the atom. The extension of our engineering technology to the very small and to the really giant belongs to the future. Power plants a mile high, or complex radio circuits the size of a pinpoint, would represent enormous new possibilities for technology.

Even within our present technology our scaling arguments are important. If we design a new large object on the basis of a small one, we are warned that new effects too small to detect on our scale may enter and even become the most important things to consider. We cannot just scale up and down blindly, geometrically; but by scaling in the light of physical reasoning we can sometimes foresee what changes will occur. In this way we can employ scaling in intelligent airplane design, for example, and not arrive at a jet transport that looks like a bee—and won't fly.

QUESTIONS

1. The total thickness of 500 sheets of a particular sample of paper is 40 mm. Assuming that they are of uniform thickness, plot a graph of the total thickness against the number of sheets. From your graph determine the thickness of 200 sheets. Of 300 sheets. Of 600 sheets.

2. A book is to have a total of 600 pages or 300 sheets (2 pages to a sheet).

   (a) Write an expression for the total thickness of the book in terms of the thickness of one sheet.

   (b) Plot the thickness of the book versus the thickness of each sheet, assuming that the sheets can be any thickness from 0.060 mm to 0.120 mm.

   (c) Actually, paper is available only in certain specified thicknesses. If we assume that these thicknesses are multiples of 0.005 mm between the limits given in (b), can we connect the points of our graph with a smooth line? Why?

3. The temperature of a container of water is read every 0.1 min as it is heated, with the results shown in the following table.

TIME *(min)*	**TEMPERATURE** *(°C)*
0.0	29.4
0.1	30.7
0.2	31.9
0.3	33.2
0.4	34.5
0.5	35.8
0.6	37.2
0.7	38.3
0.8	40.0
0.9	40.8
1.0	42.1
1.1	43.4

   (a) Plot a graph of the temperature as a function of the time.

   (b) One of the readings seems to be wrong. What temperature do you think is correct for the corresponding time?

   (c) What would you expect the temperature to be at the end of 1.2 min?

4. (a) The temperature at which a liquid boils depends upon the air pressure acting on its surface. Plot the pressures and the corresponding boiling temperatures of water from the data in the following table.

   (One atmosphere is the pressure exerted by the air at sea level.)

PRESSURE *(atmospheres)*	**TEMP.** *(°C)*	**PRESSURE** *(atmospheres)*	**TEMP.** *(°C)*
6.05×10^{-3}	0	3.12×10^{-1}	70
1.21×10^{-2}	10	4.67×10^{-1}	80
2.30×10^{-2}	20	6.92×10^{-1}	90
4.20×10^{-2}	30	1.00	100
7.30×10^{-2}	40	1.43	110
1.22×10^{-1}	50	1.96	120
1.97×10^{-1}	60	2.67	130

   (b) Notice that your graph does not present the data for low pressures clearly. Make a special plot for temperatures up to 60°. What do you observe?

   (c) The starch in potatoes will not convert to the edible form at temperatures of less than about 90° C. What is the lowest pressure at which potatoes may be cooked by boiling in an open pan?

5.* If $W = kV$, what is the effect on W of tripling V?

6.* If $W = kV$, what does a graph of W as a function of V look like?

7. A piece of copper wire 60 cm long and 0.070 cm in diameter was loaded with varying masses, and the stretch of the wire was determined. The data obtained are given in the table:

LOAD	**STRETCH**
(kg)	*(cm)*
0.0	0.0×10^{-2}
0.5	1.0×10^{-2}
1.0	2.2×10^{-2}
1.5	3.2×10^{-2}
2.0	4.2×10^{-2}
2.5	5.3×10^{-2}
3.0	6.3×10^{-2}
3.5	7.3×10^{-2}
4.0	8.3×10^{-2}
4.5	9.7×10^{-2}
5.0	11.3×10^{-2}
5.5	15.3×10^{-2}

(a) Plot a graph of the stretch of the wire as a function of the load.

(b) Over what range is there a direct proportion between the stretch and the load? What is the value of the proportionality constant in this range?

(c) What would the stretch have been if a load of 3.3 kg had been applied? What would you predict for a load of 20 kg?

8.* The total surface area of a rectangular solid is the sum of the areas of the six faces. If each dimension of a given rectangular solid is doubled, what effect does it have on the total surface area?

9.* How many state maps of scale 1:1,000,000 would you need to paper the state with those maps?

10. The volume of one cylinder is eight times the volume of another similar cylinder.

(a) If the circumference of the base of the smaller cylinder is 0.5 m, what is the circumference of the base of the larger cylinder?

(b) If another cylinder, also having eight times the volume, has the same height as the smaller cylinder, what is the circumference of its base?

11. A hollow metal sphere has a wall thickness of 2 cm. If you increase both the diameter and the thickness of this sphere by the same factor so that the total volume is three times the original volume, how thick will the shell of the new sphere be?

12.* What will be the area of a square surface constructed in the pyramid shown in Fig. 2 – 14, if it is 10 m from point *P*?

13. If two identical lamps are 60 cm apart, where should a screen be placed between them so that the intensity on one side of the screen is four times the intensity on the other side?

14. If you start at any time with a given amount of polonium, one-half of it will remain at the end of 138 days. With this information, make a graph from which you can predict how much of a sample of radioactive polonium you have today will remain after any number of days up to a year. About what percent of a given sample of polonium will remain after one year?

15. Consider a one-celled animal which reproduces by binary fission (one cell divides and becomes two) on an average of every 30 minutes when food and other conditions are right. Assume that each cell reaches a diameter of about 10^{-5} meters before it divides.

(a) Make a graph of the total population as a function of time, starting with one individual.

(b) Write an expression for the total population after n periods of 30 minutes each.

(c) Can you extrapolate the graph back to the point where the population is zero? Can you interpolate the graph between the points?

(d) Can you extrapolate the graph ahead for 50 hours? Would such an extrapolation be meaningful? (Hint: calculate the number of cells and the volume they would occupy.)

16. (a) Using tables of natural logarithms and logarithms to the base 10, plot on the same graph both log *y* and ln *y* as a function of *x* using the values of *x* and *y* given in the data below.

x	*y*
1.0	0.5
4.0	2.0
6.0	18.0

(b) What is the ratio of the slope $\dfrac{\Delta \log y}{\Delta \log x}$ to the slope $\dfrac{\Delta \ln y}{\Delta \ln x}$?

(c) What is the expression relating *y* and *x*?

17. A quantity *y* of a radioactive element decays exponentially according to the equation $y = y_0 e^{-kt}$. Show that if in time t_1 the quantity is reduced to half its value at $t = o$, it will be reduced to a quarter of its original value in time $t_2 = 2t_1$.

18. Plot a log-log graph of the data in Table 2.

19. The logarithmic form of the general equation for the power law, $y = kx^n$, is $\log y = \log k + n \log x$,

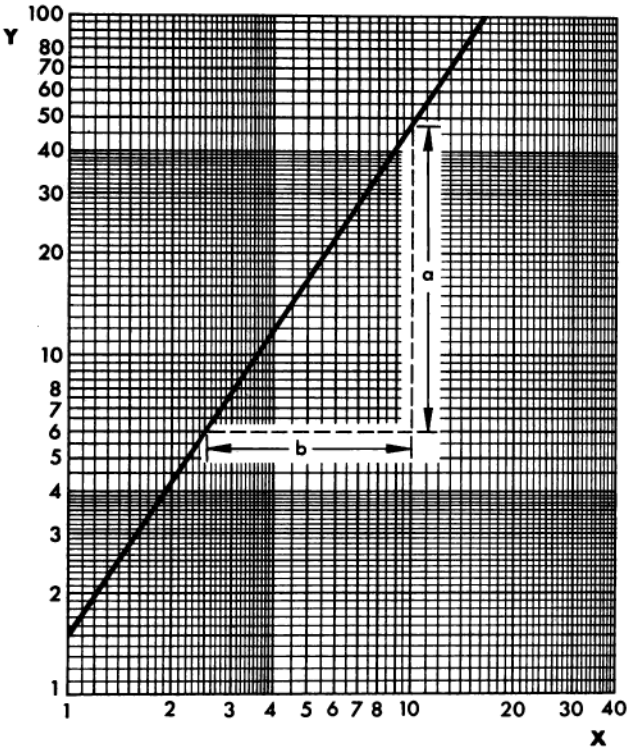
where log k is the value of log y when $x = 1$ which means that log $x = 0$).
This expression gives

$$n = \frac{\log y - \log k}{\log x}$$

Use this expression to find the value of n in Fig. 2 – 27.

(a) How does the value you obtain compare with the slope calculated by finding the ratio of $\frac{a}{b}$ when a and b are measured in centimeters?

(b) What is the mathematical expression relating y and x?



2–27 For Question 19.

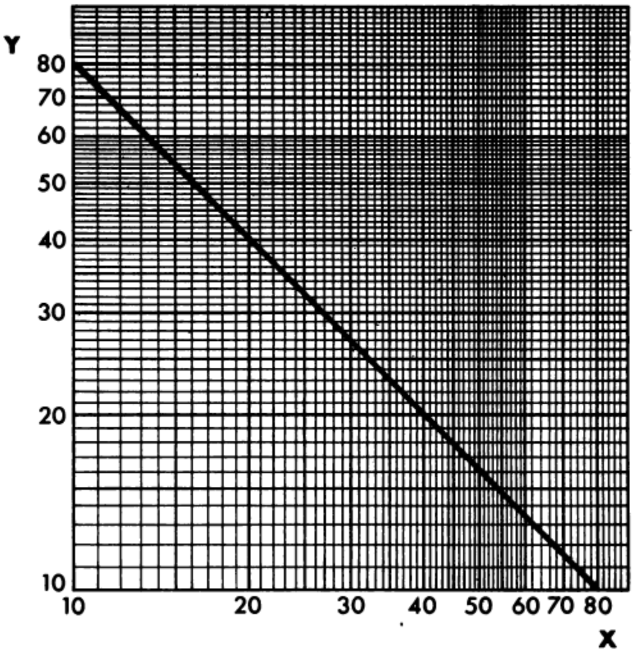
20. The table below shows the period of revolution of several planets of the solar system and the mean radii of their orbits.

	MEAN RADIUS, R, OF ORBIT (meters)	PERIOD OF REVOLUTION, T (seconds)
Venus	1.08×10^{11}	1.94×10^{7}
Earth	1.49×10^{11}	3.16×10^{7}
Mars	2.28×10^{11}	5.94×10^{7}
Jupiter	7.78×10^{11}	3.74×10^{8}
Saturn	1.43×10^{12}	9.30×10^{8}

(a) Plot the period of revolution as a function of the mean radius of orbit for the planets listed in the table. What type of function is suggested by your graph?

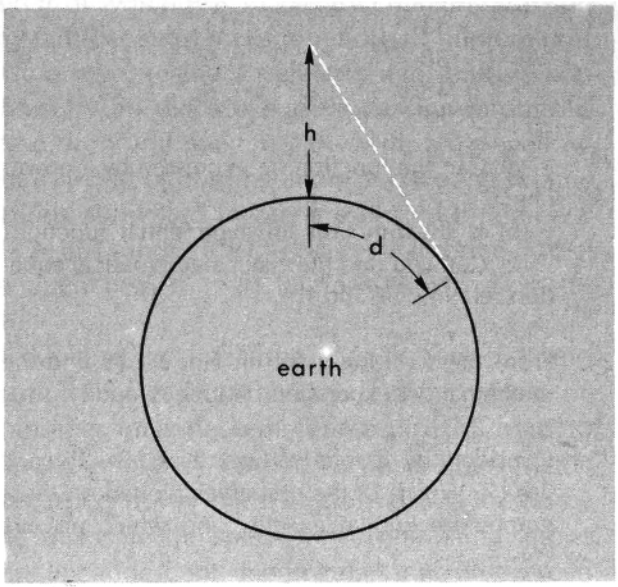
(b) Plot the same data on log-log paper. What can you conclude about the relation between T and R?

21. Find the mathematical expression relating y and x that is plotted in Fig. 2 – 28.



2–28 For Question 21.

22. The data at the top of the next page relate the altitude h of a point above the surface of the earth to the distance d along the surface from a spot directly below the point to the horizon as viewed from the point at an altitude h. (Fig. 2 – 29.)



2–29 For Question 22.

(a) Over what range can the function relating h and d be accurately expressed as a power relating h and d?

(b) What is the form of this power law?

(c) What percent error would result if you used this power law to calculate the horizon distance for $h = 1.00 \times 10^7$m?

LOG H (m)	**LOG D** (km)
0.50	0.78
2.00	1.52
3.00	2.02
5.00	3.02
6.00	3.51
7.00	3.87
8.00	3.98
9.00	4.00

23. The data below show the results of an experiment in which the temperature of a container of warm water was measured as it slowly cooled to room temperature. The difference ΔT between the water temperature and room temperature is recorded as a function of the cooling time t.

TIME, t (min)	**TEMPERATURE DIFFERENCE, ΔT** (°C)
2	37
4	34
6	31
8	29
13	23
18	19
24	16
28	13
33	11
38	9
48	6

(a) Can the function be expressed by a power law?

(b) Is the function an exponential function?

(c) Can you find the exact mathematical equation relating ΔT and t?

24. From the semi-log graph in Fig. 2 - 19 find the mathematical expression relating N and t.

25. Consider the graphs of Figs. 2 - 6, 2 - 16, and 2 - 17. Which of the quantities plotted on these graphs are known exactly, and which are not known exactly because they are measured quantities?

26. Use the graph you made for Question 7 to construct a graph of the stretch as a function of load for

(a) a copper wire 120 cm long and 0.070 cm in diameter.

(b) a copper wire 60 cm long and 0.035 cm in diameter.

27.* Suppose you scale up all the linear dimensions of a steel wire by a factor of three. By what factor does each of the following change?

(a) Volume
(b) Mass
(c) Surface area
(d) Breaking strength

28. The leg bones of one animal are twice as strong as those of another, closely related animal of similar shape.

(a) What would you expect to be the ratio of these animals' heights?

(b) What would you expect to be the ratio of their weights?

29. According to the curator of the zoo, an elephant of mass 4.0×10^3 kg consumes 3.4×10^2 times as much food as a guinea pig of mass 0.70 kg. They are both warm-blooded, plant-eating, similarly shaped animals. Find the ratio of their surface areas, which is approximately the ratio of their heat losses, and compare it with the known ratio of food consumed.

30. Horror movies sometimes present a creature such as a spider large enough to devour a human being. Is such a creature a possibility? Why?

31. A sphere of iron is suspended from a wire, 0.1 cm in diameter, which is just strong enough to support it. What must be the diameter of the wire needed to support a similar iron sphere having

(a) five times the original volume?
(b) five times the original diameter?

32. A rectangular water tank is supported above the ground by four pillars 5 meters long whose diameters are 20 cm. If the tank were made ten times longer, wider, and deeper, what diameter pillars would be needed? How much more water would the tank hold? Neglect the weight of its pillars.

3

The Propagation of Light

MOST OF the information that reaches us comes to us through the sense of sight. From the very beginning of history, men have puzzled about the nature of the light that affects our eyes. The questions they asked were probably the same that have occurred to you. What is light? How does it travel, and how fast? Is seeing always believing? Why are some objects colored, some white, and some black?

3–1 SOURCES OF LIGHT

The sun, the stars, lamps, even lightning bugs, give off light. They are called *luminous* bodies (from the Latin word *lumen,* meaning light). Other objects—trees, grass, the pages of this book, for example—are nonluminous. They are visible only when they receive light from some luminous source and reflect it to our eyes.

Whether a body is luminous or nonluminous depends as much on its condition as on the material of which it is made. By changing the conditions we can make many familiar substances luminous or nonluminous at will. The filament, or fine wire, inside an electric light bulb is nonluminous unless it is heated by an electric current passing through it. We can take a cold piece of iron and make it glow with a red, yellow, or white light by heating it in a bed of burning coals or over a gas flame. When solids and such liquids as melted metals are heated to temperatures above 800 degrees Celsius (about 1400 degrees Fahrenheit), they become sources of light. Such heated materials are known as *incandescent* bodies.

Careful observation shows that the light from candle flames comes from many small, hot particles of carbon that are burning within the flame. Thus the flame is another incandescent source of light. Many of the carbon particles are not completely burned in an ordinary flame. They cool off as they are carried up by the rising air above the flame, become nonluminous, and make up the main part of the rising smoke.

Not all light sources are incandescent. Neon tubes and fluorescent lamps, like electric light bulbs, give off a bright light as long as an electric current passes through them; touching each of them, however, convinces us at once that there must be a difference in the way the light is produced. The neon and fluorescent tubes remain quite cool, whereas the incandescent bulb soon becomes too hot to touch. Pursuing this difference further, we find that by gradually increasing the current in the filament of the incandescent bulb we can increase its brightness, but that there is also an accompanying change in color. At first we see a dull red glow which changes to a bright yellow and, with sufficient current, can become what we call "white hot" like the heated iron. On the other hand, if we increase the current through a neon tube to increase the brightness, we observe no change in color. Thus we have discovered a basic difference between incandescent and other sources of light. In the former, changes in brightness, temperature, and color seem to be closely linked, while in the latter the color of the source depends mainly on the nature of the material and does not vary with brightness.

A great deal of light reaches our eyes from nonluminous surfaces. To convince ourselves, we need only to imagine how the average room would appear if the walls and other surfaces were covered with a paint so black that it reflected none of the light reaching it. The lights would appear as bright glares against dark backgrounds. If the earth had no atmosphere to reflect or scatter sunlight, the sun and the stars could be seen at the same time against a dead-black sky. White ceilings or bright walls reflect and diffuse much of the light they receive and so increase the brightness within the room. In fact, when we use indirect lighting we hide the lamps from sight, and all of the light reaches us after being diffused from the walls and ceiling. On a larger scale the moon, which we often think of as a source of light at night, is really an indirect-lighting device that reflects sunlight.

3-2 TRANSPARENT, COLORED, AND OPAQUE MATERIALS

When you look through a clean window at a brightly illuminated scene outside, you are hardly conscious of the fact that the glass is there. Substances that transmit light in this way are said to be *transparent*. Later in the day, when dusk has come, look through the same window from inside a lighted room. In addition to the world outside, you now see in the glass a reflection of yourself and of the room. The light by which you see yourself must have started from within the room. Instead of going through the glass to the outside, this light was returned to you. It was reflected.

Does the thickness of a transparent body have any effect on the amount of light it transmits? A single piece of the glass seems to transmit light almost perfectly, but if you pile up a thick stack of ten or twenty pieces of clear glass, some light is *absorbed* and the light passing through it dims and appears somewhat colored. Evidently we become aware of clear materials like plastic, glass, and water partly because they reflect as well as transmit light, and partly because some of the light is absorbed.

Thus far, we have dealt only with colorless, almost perfectly transparent bodies. We all know that there are many transparent materials which are colored. We can see objects through them, but the objects appear in different colors. These materials clearly do something to the light passing through them. Do they add something to it, or do they take something away? At first glance this may seem like a very difficult question, yet a few simple experiments can give us the answer.

First of all, look through a piece of red glass at a sheet of white paper illuminated by sunlight or by electric light. The paper now appears red. White light left the paper, but the red glass did something to the light to make it appear red. Suppose a second piece of red glass is placed between the first piece and your eye, so that the light must pass through both pieces. If the glass adds something to the light, the paper should appear a more brilliant red than before. If the glass subtracts something, we might expect that most of this "something" would be removed by the first piece of glass. The second piece should then have very little effect on the appearance of the paper. In fact, if we actually try two pieces of red glass this way, we find that the second has little effect.

We can account for the little change in color by arguing that one piece does not subtract quite all of the nonred parts of the light; there is something left for the second red glass to take away, and the red color we see through the two pieces may appear somewhat redder. We can do the same experiment with two pieces of green glass just as we did with the two pieces of red glass. As might be expected, the second green plate has little effect; at most, the paper appears a bit more green. Now what will happen if we try red and green together? After subtracting *everything but green* with the green glass, and *everything but red* with the red glass, we can certainly expect that what is left will be neither red nor green. At this stage of our investigation, we have not yet enough knowledge of the process of subtraction to predict what the color will be, but we can certainly say that much less light should reach us through the two pieces of glass than through either alone. When we try the experiment, we find, surely enough, that the brightness of the paper is tremendously reduced and that the little color left is neither red nor green but is rather a weak yellow or amber. This result convinces us of the correctness of the subtraction idea.

What is the "something" that is subtracted from white light to give color? This is a question whose answer we must postpone until we have learned a good deal more about light. Perhaps we can be satisfied for the moment with the recognition that white light is something more complex than light of a single color. We shall not, in fact, come anywhere near to complete answers to many of the questions about color in this book. As any artist will tell you, the color you see is determined by the nature of the illuminated object, the nature of the light that is illuminating it, and the condition of your eye at the time that the color is seen. It also depends on the other objects you see at the same time.

Red and green glass together subtract almost everything from white light. Most common materials go further and either reflect or absorb everything in white light, allowing no light to be transmitted. We cannot see through thin pieces of metals, wood, cardboard, or heavy cloth. Such materials are *opaque*. Opaque materials reflect some of the light that falls on them and absorb the rest, so that no light passes through them.

3–3 REFLECTION

All bodies, whether transparent or opaque, reflect some of the light that falls upon them. That is, they send some of the light back to the same side of the body as that from which it came. Most surfaces give a *diffuse reflection*. They send light off in many directions. It is by the help of this diffused light that we see illuminated bodies, observe their texture and color, and distinguish them from their surroundings.

A few materials, such as highly polished sheets of silver, aluminum, or steel, absorb little from white light and also reflect in a much more regular manner than do rougher surfaces.

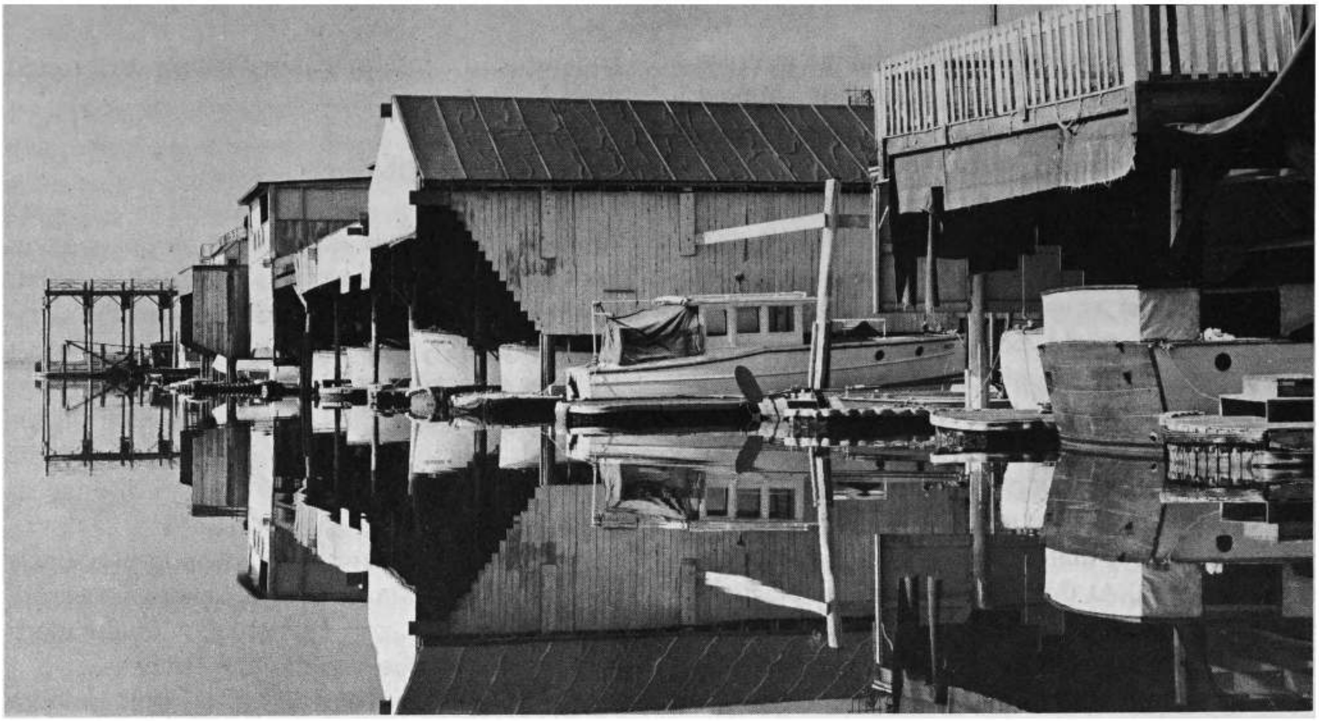
It is the regularity of reflection from smooth surfaces that allows the formation of images. Figure 3 – 1 shows a waterfront scene with boats and piers reflected in the water. Is the picture printed right-side up? Turn the page upside down and examine it again. There are several clues in the picture that tell you if it is right-side up or not.

What can you learn about images by simple observation? Look at yourself in a mirror. Is the view that you get of your image exactly like the view that other people get of you? Move your right hand. Does your image move a right or a left hand in response? Place an object such as a pencil against the mirror, then hold a similar one in your hand and observe the images. Does the image of the object in your hand seem to be on the mirror beside the other pencil, or behind the mirror? How do the pencils compare in apparent size? How does the distance of the image of the object behind the mirror compare with the distance between the pencil in your hand and the mirror?

In Chapter 4 we shall try to find out something about the laws of reflection that will allow us to discuss "why" as well as "what" happens when images are formed by mirrors or other smooth reflecting surfaces.

3–4 LIGHT-SENSITIVE DEVICES

We have not yet mentioned how we know that light is present or what its color is. We have taken



3–1 Nearly perfect reflection. Is the picture right-side up? Which are the boats? Which are their reflections?

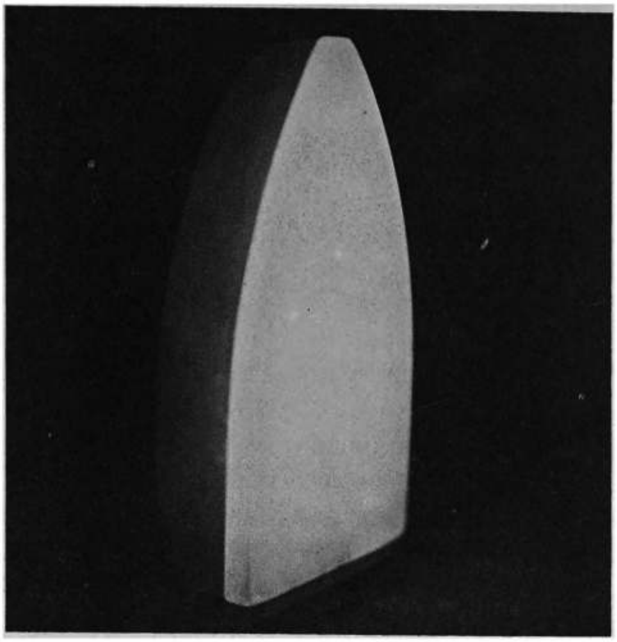
for granted that marvelous instrument, the human eye. A thorough study of the eye would include discussions of how light is refracted to fall on the nerve cells of the retina at the back of the eyeball, how the eye adjusts itself to see objects clearly at different distances or of different brightnesses, how we think the effects of color are produced, and also a discussion of the things that can go wrong with our vision. Such a study could easily occupy us for a full year or more. Here we shall have to be content with the observation that as light passes through structures in the front of the eye, it is *refracted,* which means that its path is bent in such a way that an image is formed on the retina. Chemical changes take place at the retina, and as a result electrical impulses are sent along the nerves to the brain.

There are many instruments, other than the eye, that respond chemically or electrically to the action of light. These instruments are often used in the laboratory to study light by methods that are more convenient or more satisfactory than visual observation. The best-known substances that are *photosensitive* are some of the chemical compounds containing silver. Such compounds are used in photographic films. The full process involved in the exposure and development of the film is complex, but the essential thing is that those portions of the film on which light has fallen are left, after chemical treatment, with a deposit of finely divided silver. In this state the silver does not appear as bright, shiny metal, but as dull, opaque particles, the familiar black portions of a film negative. The portions not exposed to light become clear and transparent. Color film, sensitive to different colors of light, is made by the addition of various dyes and by changes in composition.

There are also a number of devices that produce an electric current when light falls on them, without any chemical changes taking place. Chief among these are photoelectric cells, or photocells, which we shall discuss later. The electric current that flows through a photocell is proportional to the light intensity.

3–5 INVISIBLE LIGHT

Photocells, photographic films, or other detectors of light sometimes show the presence of light when none is visible to the eye. For example, Fig. 3 – 2 shows a photograph of an ordinary flatiron, taken in a totally dark room using a special kind of photographic film. Apparently the flatiron is giving off "light" of some sort, although we do not see it as a luminous object. We might guess that the light given off is the same thing as the heat which we feel when we hold a hand a few inches away from a heated iron. As we shall see



3–2 Photograph of a flatiron taken by its own invisible "light." What is the name of this invisible "light"?

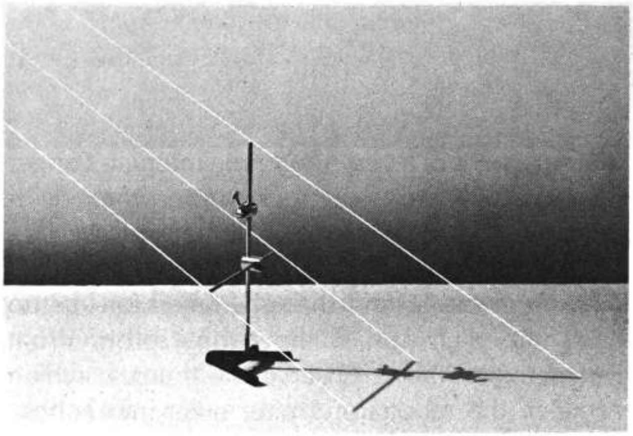
later, this guess is correct: *heat* radiation is indeed a form of light known as *infrared* radiation. About one-third of the radiation that reaches us from the sun is heat radiation; most of the rest is visible light.

Like the eye's retina, an ordinary photocell cannot detect heat radiation. This may be shown by placing a photocell near a heated flatiron in a dark room. We know from the photographic evidence that the iron gives off heat radiation, yet the photocell shows no response. A different result is obtained, however, if we use a mercury-arc lamp as a source of light. A good lamp for this purpose is one of the small sterilizing lamps sometimes used in refrigerators. (These lamps are constructed with a special kind of glass, known as Corex.) When a photocell is placed near such a lamp, the electric current passing through the cell increases with the intensity of the light striking it. If a piece of ordinary window glass, which is almost perfectly transparent to visible light, is placed between the lamp and the photocell, the current in the photocell decreases appreciably. From this simple procedure we learn several things. We observe no difference in the light from the lamp when looking through the window glass, but the photocell obviously does. There must be some part of the radiation from the lamp which is invisible to the eye and which is not transmitted by window glass. It is evident, too, that Corex must be transparent to this part of the radiation; otherwise it would not reach the photocell. As we shall see, the mercury lamp emits *ultraviolet* light in addition to visible light. This radiation is not only invisible—it is harmful to the human eye. One should therefore avoid looking at sources of ultraviolet light.

3–6 HOW LIGHT TRAVELS

The sun and the stars are so familiar that we seldom think of the vast stretches of almost empty space that separate them from us. Yet we know that the sun is about 1.5×10^{11} meters away from the earth, that the nearest star is about three hundred thousand times as far, and that countless stars have been seen at distances so remote that comprehension almost fails us. All the information from which we have learned about this vast universe has come to us "riding swiftly astride beams of light." It must therefore be true that light can travel over very great distances and that it can travel freely in space.

As familiar as the sun itself are the shadows it casts. What can we learn from them about light? As we walk or run along on a sunny day, our shadows keep pace with us. This simple experience shows that light must travel much faster than we can run, for the shadows of our heads must lag behind us by the distance we move while light goes from our heads to the ground. Shadow shapes, too, reveal information about the nature of light. In Fig. 3 – 3 a set of



3–3 The formation of shadows. Lines connecting points on the shadow to corresponding points on the object casting the shadow are nearly parallel; all lines point to the light source.

lines connects various points on a shadow to the corresponding points on the object which casts the shadow. The lines are almost parallel and they all point to the source of the light. Evidently light travels in straight lines.

3–7 THE SPEED OF LIGHT

All of us have, at one time or another, heard the roar of a jet plane high in the sky and instinctively looked for it in the direction of the sound. We finally saw it far ahead of the sound. Then in deciding where the airplane actually was, we believed our eyes rather than our ears. Why did we trust our eyes? We know that sound takes some time to travel so great a distance and that light travels much faster than sound. We did not assume that the plane was beyond the point where we saw it, because we believe that light travels fast indeed!

Galileo suggested a method for finding the speed of light, similar to the method he used to measure the speed of sound. Two men with lanterns were placed at a measured distance apart. The first uncovered his lantern and started a clock, and the second uncovered his lantern when he saw the light from the first. When the first man saw the second's light, he stopped the clock, thus measuring the time for light to travel from the first man to the second and back—or so Galileo hoped. This experiment failed to give the speed of light because light travels so fast. But it was not a complete failure. It showed that the speed of light was too great to allow its passage to be measured over short distances with the crude timing mechanisms then available.

A satisfactory method of measuring the speed of light was first proposed by Roemer in 1676. This method does not need a highly precise measurement of a very short time interval. Longer time intervals are easier to measure. Roemer increased the time interval by increasing the distance the light travels. He used astronomical measurements to find the time taken for light to cross the diameter of the earth's orbit. What Roemer did was to observe the times at which some of the moons of Jupiter went into eclipse as they passed into that planet's shadow. These eclipses occur at regular intervals. For example, the brightest of the satellites revolves in its orbit about Jupiter in just a little over seven days and it is eclipsed once during every revolution. Using the average time between eclipses as the time of rotation of the satellite, he found that the eclipses occurred about 11 minutes early when the earth was nearest Jupiter and 11 minutes late when it was farthest away, because the light had to travel across the earth's orbit. Roemer could not find a good value of the speed of light from these observations. His measurements were inaccurate and also, at that time, the size of the earth's orbit was not accurately known. The great contribution of Roemer's work was the demonstration that light took a measureable time to cross the earth's orbit and that it therefore travels at a finite speed.

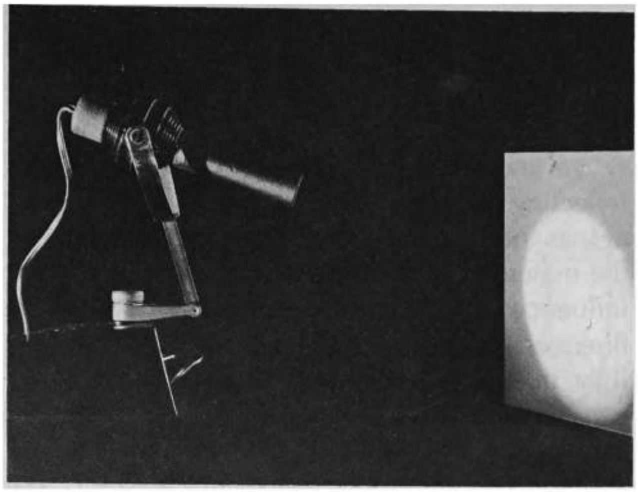
Later measurements of the delays in eclipse times showed that the interval required to cross the orbit is 16 min 20 sec rather than 22 min as Roemer found. The average distance of the earth from the sun is now known to be 1.47×10^{11} meters; hence the speed of light is

$$c = \frac{2 \times 1.47 \times 10^{11}\,\text{m}}{980\ \text{sec}} = 3.00 \times 10^{8}\ \text{m/sec}.$$

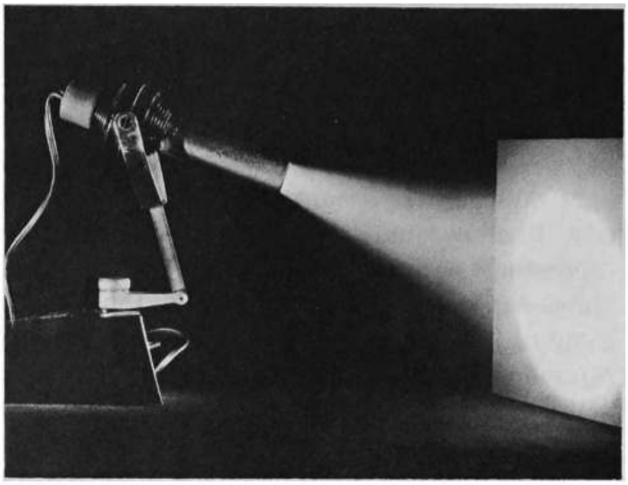
The first determination of the speed of light over a distance short enough to be practical on the surface of the earth was made by Armand Fizeau in 1848. It required the invention of a timing device that could measure very short intervals with accuracy. We shall discuss Fizeau's method in Chapter 5.

3–8 LIGHT BEAMS, PENCILS, AND RAYS

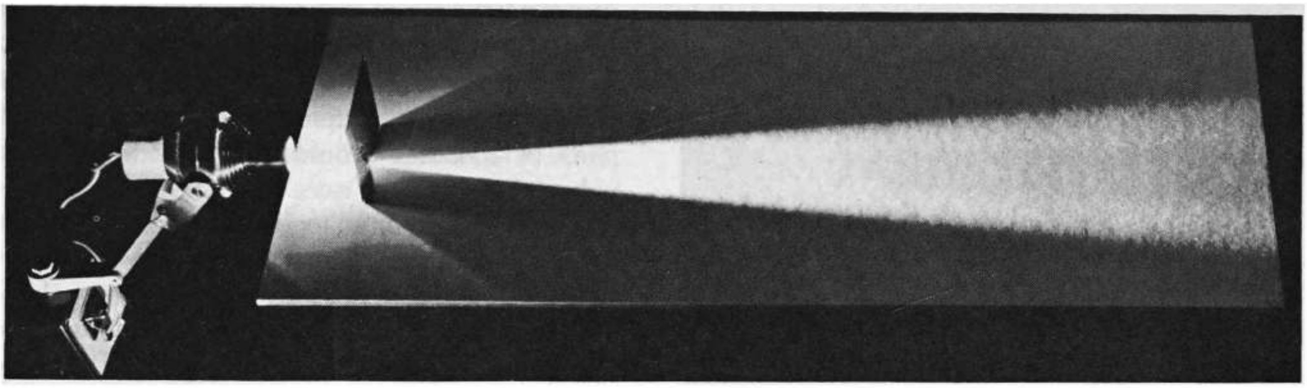
Figure 3 – 4 shows a light source which is pointed at a piece of paper. We know that the light comes from the source and we see it reflected from the paper, but we do not see it between the source and the object. In Fig. 3 – 5 fine smoke particles have been introduced into the air, and now the beam shows clearly all the way. The straight edges of the beam connect the source with the edges of the illuminated spot, confirming our belief that light travels in straight lines. The fact that we saw nothing of the beam until the smoke was introduced tells us that light enters our eyes only when we are looking directly toward a source or when there are illuminated bodies that can reflect light directly to our eyes. In this

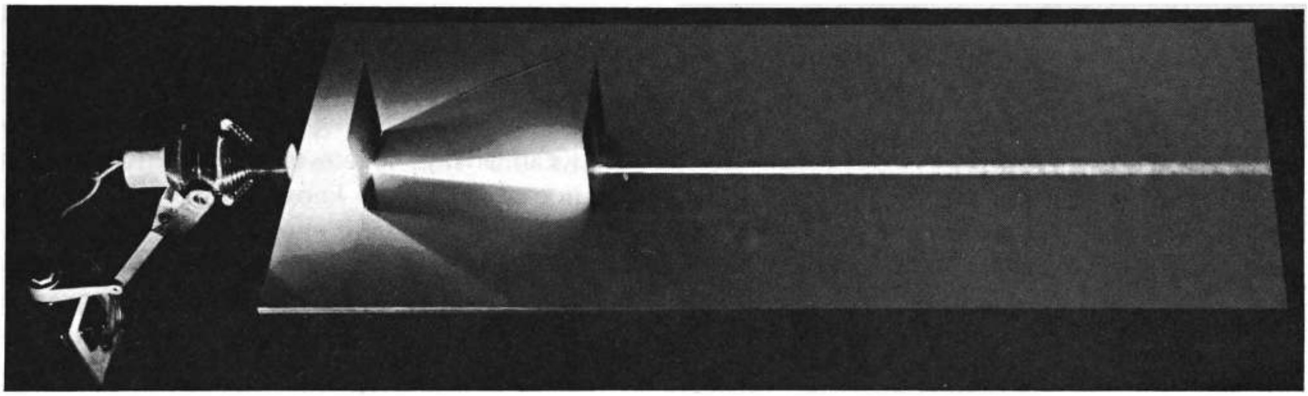


3–4 **Light beams are invisible. Here we see the light source and the reflection from the white cardboard, but nothing in between source and target.**



3–5 **If fine particles are introduced into the air, as shown above, the light beam is seen clearly all the way from the source to the cardboard target. Compare with Fig. 3–4.**





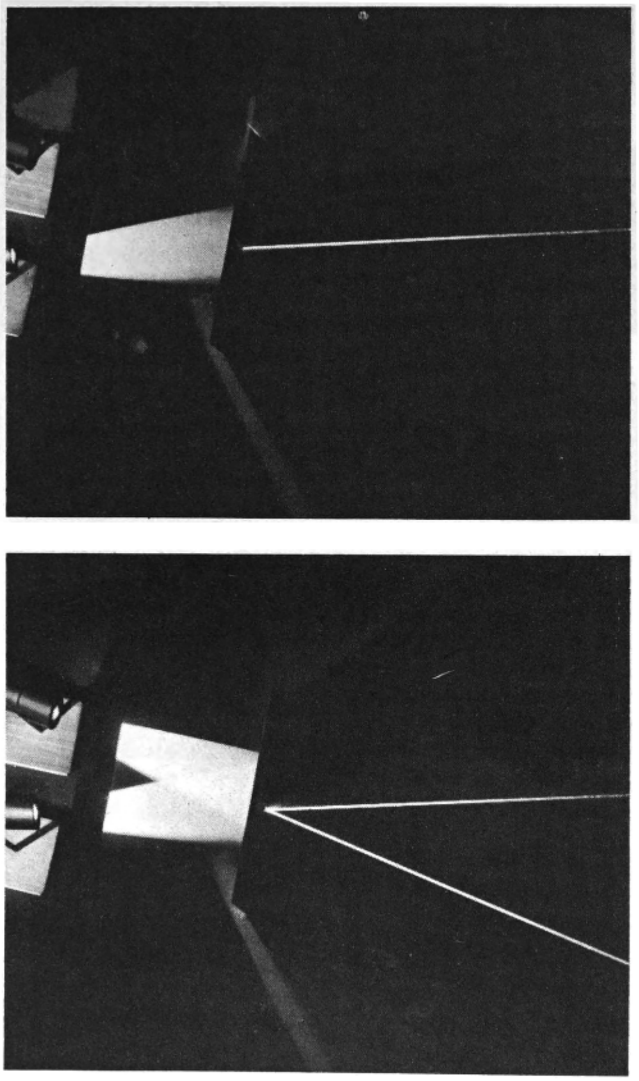
3–6 **Beams and pencils of light. In the top picture, light coming from the point source through a pinhole forms a cone-shaped beam. Bottom, a second pinhole in line with the first produces a narrow pencil of light.**

case the light was reflected from the paper and from smoke particles, making them act like new sources to send the light on new straight-line paths toward our eyes.

Figure 3 – 6 (top) shows beams of light coming from a point source through a pinhole. In the upper picture, we see a cone marked out by the beam. Can you explain why it is a cone rather than a cylinder? The axis of the cone (the line joining the center of the pinhole to the center of the base of the cone) passes through the center of the source and the pinhole.

By putting two pinholes in line, as shown in Fig. 3 – 6 (bottom), we get a very narrow *pencil* of light. It is often convenient to imagine such a pencil of light that is smaller and finer than any pencil that we can produce in practice. The finest pencil that one can imagine is just a straight line. We call this extreme pencil a *ray* of light. Of course, we can never produce single light rays, but the idea of a ray is a very useful one. It allows us to draw on paper lines that represent the directions in which light is traveling. A pencil of light can be found in nature; a ray of light is something we have invented to represent a very small pencil.

We often draw a few rays, for example, to indicate the path of the light that forms the shadow in Fig. 3 – 3. The rays we draw are not distinguishable in the light, which apparently goes in all directions from all points on the surface of the source. The rays we draw are useful like the lines on an architect's plan, but they are no more light than his lines are walls or windows. Light does not make rays; we do. They help us describe the way light behaves.



3–7 An experiment showing that two beams of light can pass through each other.

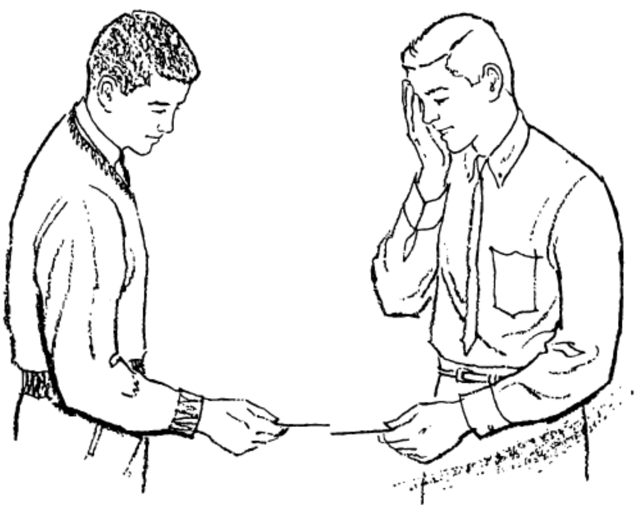
We know that light beams or pencils striking reflecting objects illuminate them so that they act as new sources. Light thus interacts with the material on which it falls, i.e., the light beam influences the material, and the material influences the behavior of the light beam. Can light beams also interact with each other? We can answer this question by letting the light beams from two different sources pass through a single pinhole, as is shown in Fig. 3 – 7. We find that each beam passes through the hole as if the other beam were not there.

In general, each light beam from two or more sources behaves as if the other beams were not present, and this independence of action is very important. The procedure of studying the pencils or rays from different sources or different parts of a source would otherwise be of little value. Tracing the rays that show the directions of two jets of water, for example, does not tell us much about where they go when they hit each other. It is because the light from each of two sources acts as if it were alone that we can trace rays to find the regions of light and shadow when both sources are present.

3–9 HOW WE LOCATE OBJECTS

We can always locate the position of a point source of light if we know the direction of several of the rays that come from this source. We simply draw two or more of these rays backward until they meet. The point of intersection is the location of the source. When a cone of rays comes to our eye from such a source, we automatically change the shape of the eye so that the diverging rays from the source are focused, and we can see an image of the source. This process of focusing our eyes gives us information equivalent to that of tracing back the rays, and in certain circumstances, without being aware of it, we use this information to estimate the distance to the source.

We can estimate distances more easily with two eyes than we can with one. To convince yourself of this, have a friend hold a thin piece



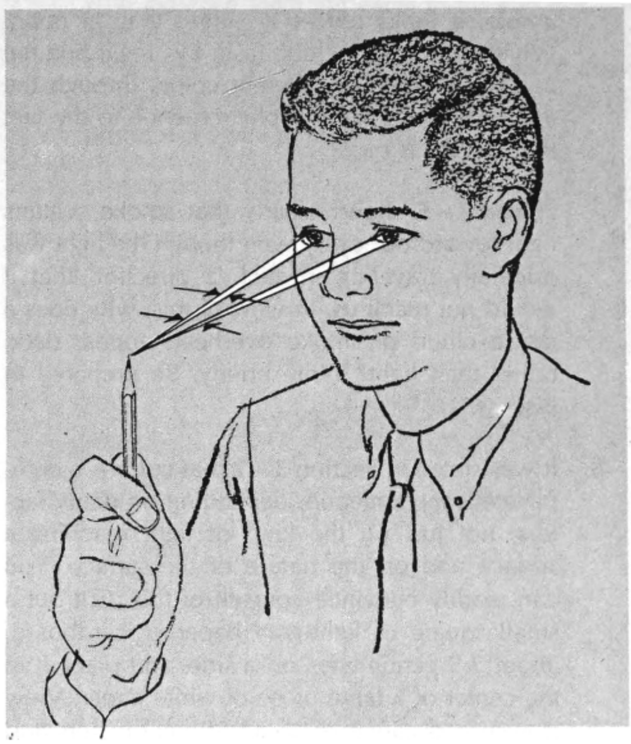
3–8 **An experiment for testing your ability to judge distance with one eye.**

of wire away from other objects (Fig. 3 – 8). With another similar wire in your hand, try to touch the end of your wire to the end of his wire when it is about as far away as you can reach. You will find that you can do this with considerable accuracy. Now try the same experiment with one eye closed. You will find it much more difficult to bring the ends of the two wires together.

One of the reasons you judge distances better with both eyes open is indicated by this experiment. Light coming from the end of an object must travel in different directions to reach the two eyes (Fig. 3 – 9). This is a simple case of triangulation (see Chapter 1). Here the base line is the distance between the eyes; and if the two eyes were independent persons, they could figure out the distance to the object by plotting the base line and the angles to the lines of sight or by doing trigonometric calculations. Of course our brain does not tell us the distance by such calculation. Instead the whole act of converging our eyes on the right sight lines and focusing them to the right cones of rays has been calibrated by our past experience. Our acquired knowledge of the distances associated with the particular act of converging and focusing tells us directly where the object is—and it tells us with more accuracy than just the act of focusing with one eye.

If you leave other objects around when you try to measure distance, or if you use blocks instead of thin wires, you will be able to touch one object to another still more easily. In such circumstances our brains take advantage of other clues, for example, the three-dimensional effect that arises from the slightly different view of the faces of the blocks seen by each eye. These two views differ more as the object is moved closer. The size and position of nearby objects also help the brain to judge the positions of the wires.

There are many clues, such as size, shadows, and motion of familiar objects, that help us to estimate how far an object is from us. When these clues are present, the direct interpretation of the rays in the cones of light entering the eyes and of the angle between the cones becomes less important, but at short distances these obvious physical clues are definitely used.



3–9 **Binocular vision. We judge distance by the angle at which the eyes are converged on the subject, as well as by focusing them on the cones of rays.**

QUESTIONS

1.* Which of these objects are luminous (when in normal operation)?

Camera
Firefly
Flash bulb
Mirror
Chrome trim on car
Electric stove heating element
Diamond

2.* How can you tell whether an object is luminous or nonluminous?

3. Try to see whether light reflected from colored glass takes on color from the glass.

4.* Under white light object *A* appears white and object *B* appears green. How will they appear under red light?

5. We have seen that glass, although transparent, does not transmit all of the light that enters it, some of the light being absorbed. Is this also true of clear water? Be prepared to discuss what evidence you would look for to support your answer.

6. The color of the sun appears to change. For example, it looks redder at sunset than at noon. What can you conclude from this regarding the passage of light of different colors through the air? Can you relate this phenomenon to the fact that the sky is blue?

7. Figure 3 – 5 shows clearly that smoke scatters light toward our eyes, even though the light was originally traveling in such a direction that it would not reach us. In view of this, why does a dense cloud of smoke overhead appear dark, rather than light? Write briefly. Be prepared to discuss.

8. It was stated in Section 3 – 2 that color is a complicated phenomenon, depending on many factors, not just on the kind of light reaching a surface and on the nature of the surface. You can readily convince yourself of this. Cut out a small square of light-gray paper or cardboard, about 2.5 centimeters on a side, and place it in the center of a large piece of white paper. View it from a distance of about 50 centimeters and under bright illumination.

   (a) Describe its color and brightness.

   (b) Move the gray square to the center of a large piece of bright-red paper and view it under the same conditions as before. Again describe its color and brightness. Are they the same as before?

   (c) Finally, move the gray square to a piece of black paper. What happens now?

9. We have seen that when light from a white object falls on a photographic film it produces a deposit of black silver after development. This results in the familiar negative image on the film. If you shine light through the negative onto another photographic film, after development what kind of an image will you have?

10. Move this book toward you, with one eye closed or covered. When the book has just reached the point at which the print blurs, have someone measure the distance of the book from your eye. Repeat the experiment with the other eye. Are the two distances approximately equal? Try this with a few people of different ages, and record the results along with their ages.

   (a) Is there a limit to the ability of the eye to adjust itself for clear vision?

   (b) Is this limit the same for all persons, or for the two eyes of any one person?

   (c) Does the limit vary, in general, with the age of the person? Compare your answers with those found by other people in the class.

11.* In Fig. 3 – 2, why does the underside of the flatiron appear brighter than the rest of it?

12.* It is the ultraviolet radiation from the sun that causes suntanning. Can you get suntanned in a glass-roofed solarium?

13.* "Light travels in a straight line." Name instances where this is not true.

14. Remembering that we see the moon by reflected light, can you show that

   (a) a moon rising in the east at midnight cannot be a full moon

   (b) a new moon cannot be seen for long after dark.

   Hint: Make diagrams showing the positions of the sun, earth, and moon at different phases of the moon.

15. In measuring large and small distances, we make the assumption that light travels in straight lines.

   (a) Describe an experiment which shows that light travels in straight lines.

   (b) Can your experiment be modified to show that in some cases it does not?

16. We sometimes see total eclipses of the sun by the moon, and sometimes annular eclipses. In the

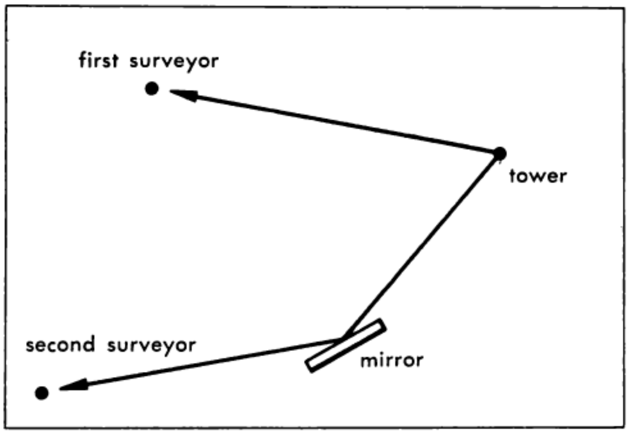
latter, a ring of light from the sun is seen around the edge of the moon.

(a) By drawing a diagram of the earth, moon, and sun, explain why two different kinds of solar eclipses occur. Do the distances of the moon from the earth and of the earth from the sun remain the same?

(b) The moon is about a quarter of a million miles from the earth, and the sun is about 93 million miles away. If the moon has a diameter of two thousand miles, what is the approximate diameter of the sun?

17. Artificial satellites can often be seen as bright objects high in the sky long after sunset. What must be the minimum altitude of a satellite moving above the earth's equator for it to be still visible directly overhead two hours after sunset?

18.* Two surveyors are determining the location of a tower by triangulation. One of them is fooled by a mirror, as shown in Fig. 3 – 10. Where will they locate the tower?



3–10 For Question 18.

19. (a) How long does it take light to reach the earth from the sun?

(b) If the light from the nearest star takes 4.3 years to reach us, how far away is the star?

(c) Why is it convenient to express distances to stars in terms of light-years, rather than in meters, kilometers, or miles?

20. Radio waves travel at the same speed as light in empty space or in air.

(a) How long does it take a radio signal to travel from New York to San Francisco, a distance of about 4.8×10^3 km?

(b) A radar transmitter, which sends out radio signals of a particular type, when pointed at the moon receives a reflection 2.7 sec after the signal is sent. What does this experiment give as the distance of the moon from the earth?

21. (a) How would the 2.7 sec required for a radio wave to travel to the moon and back affect the conversation between an astronaut on the moon and ground station personnel?

(b) Extend (a) to astronauts on Mars when it is at its nearest or farthest position from the earth. (The distance from the earth to the sun is 1.5×10^{11} m and from Mars to the sun is 2.3×10^{11} m.)

(c) Extend (a) to an astronaut traveling close to the next nearest star, from which the light requires 4.3 years to reach us.

4 Reflection and Refraction

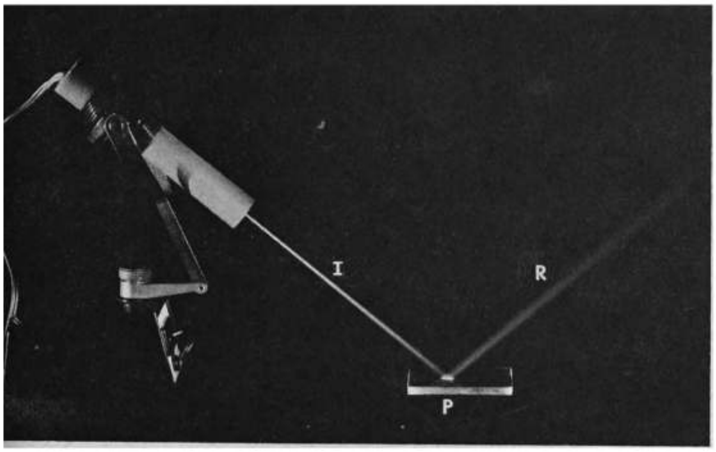
4–1 THE LAWS OF REFLECTION

All that we have said in Section 3 – 9 applies to finding the *apparent* position of a source of light. This position, however, need not be the actual position of the source. If the rays of light coming from the source have had their direction changed by reflection (see Chapter 3), when we trace the rays straight backward their point of intersection will not be the actual point from which the rays started. Nevertheless we shall see the source apparently located at the point of intersection. We are seeing an image of the source, not the source itself.

In Chapter 3 we briefly discussed the images seen in mirrors. Can we find laws of reflection that explain the position of such images? Figure 4 – 1 (a) shows a pencil of light falling on a polished metal plate; the pencil is made visible by smoke. We notice at once that the *reflected* pencil *PR* is almost as sharply defined as is the *incident* pencil *IP*. Polished metals, liquid surfaces, and mirrors that reflect in such a fashion that the reflected pencils are sharply defined are said to be *specular reflectors*. Materials such as white paper give *diffuse reflection*. [See Fig. 4 – 1 (b).]

To study specular reflection we place a sheet of white cardboard on the reflecting plate in such a way that the two pencils just skim along its surface as in Fig. 4 – 2. We find that the cardboard must be held perpendicular to the reflecting surface. If we do the experiment over and over again, changing the direction in which the incident pencil strikes the surface, we always get the same result—the cardboard (and therefore the plane of the two pencils) is always perpendicular to the reflecting surface.

As a further step draw a line through the middle of the cardboard at right angles to one edge. Then place the cardboard in a position perpendicular to the reflecting surface as before, with the end of the line just touching the point at which the incident pencil strikes. Since the line is on the cardboard, it is in the same plane as the two pencils of light and also perpendicular to the reflecting surface. This observation gives us the first law of specular reflection:



4–1a Specular reflection. A beam of light striking a polished metal surface produces a sharply defined reflected beam.

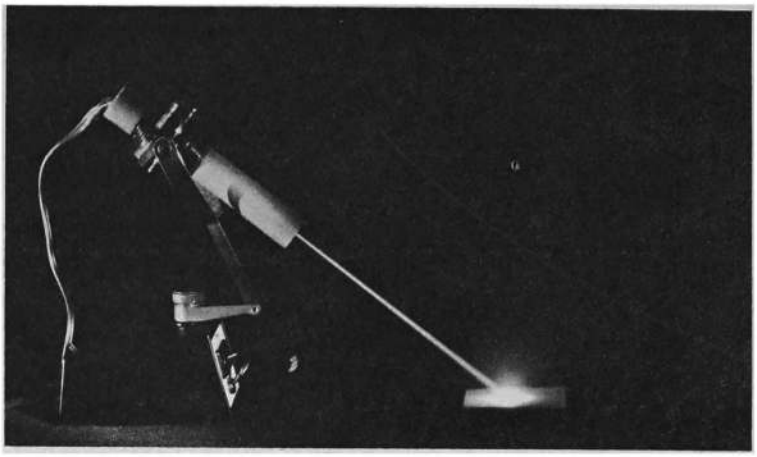
When light is reflected from a plane specular surface, the incident ray, the reflected ray, and the normal (the perpendicular) to the surface at the point of contact all lie in the same plane.

Now suppose that we measure the angles between the normal and the two rays, as indicated in Fig. 4 – 2. We find that the angle between the reflected ray and the normal, called the *angle of reflection,* is equal to the angle between the incident ray and the normal, called the *angle of incidence.* This result is confirmed whenever we compare these two angles; hence we can state the second law of specular reflection:

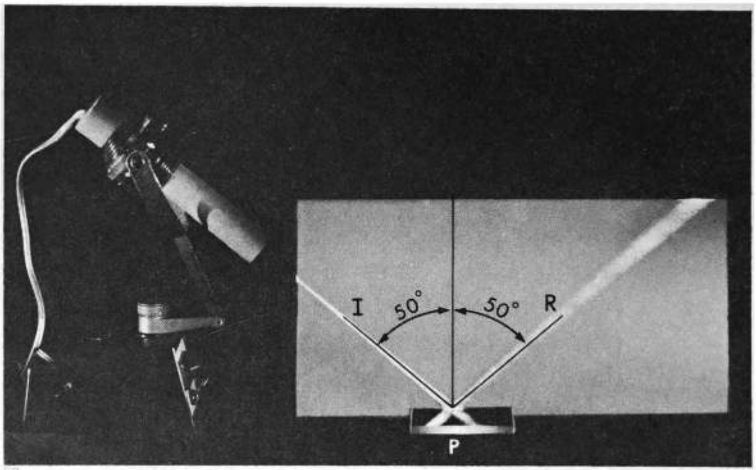
The angle of reflection is equal to the angle of incidence.

4–2 IMAGES IN PLANE MIRRORS

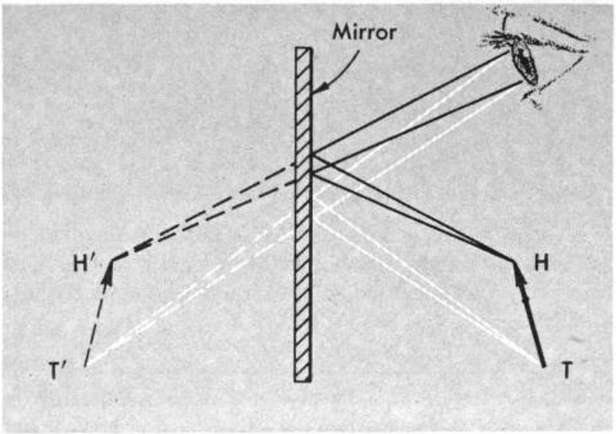
Can these two laws of reflection help us locate and explain the nature of the images in ordinary mirrors? In Fig. 4 – 3 an arrow, *HT,* is placed near a mirror. We now ask whether light rays starting from any point, such as the tip of the head of the arrow *H,* and reflected from the mirror to the eye, seem to come from a common point. The reflected rays were accurately drawn according to the laws of reflection. As the figure shows, the two rays from *H* reach the eye as if they came from the common point *H′*. This suggests that the tip of the arrowhead appears to be at *H′*. Similarly, the white rays in the figure indicate that the tail of the arrow, *T*, appears to us as if it were at *T′*.



4–1b Diffuse reflection. When a beam of light strikes a piece of white paper, light is reflected in all directions.



4–2 An experiment that demonstrates the first and second laws of specular reflection. The beams seem to cross because you see in the mirror the reflections of the lighted regions on the cardboard.



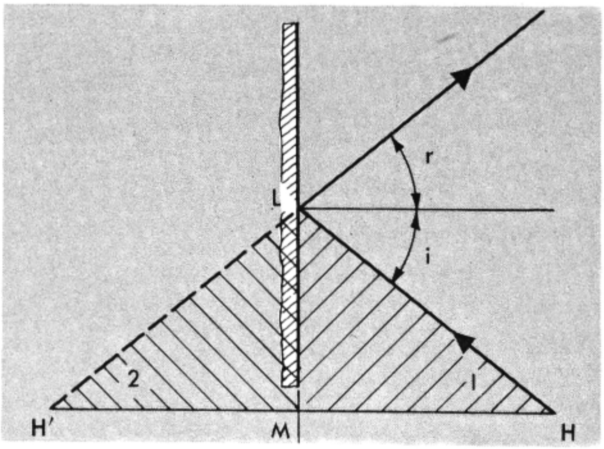
4–3 The formation of an image in a plane mirror.

If we were to draw a third ray originating from *H*, would it also appear to come from *H′*? What about a fourth ray? We shall use the laws of reflection to prove that all the rays originating from *H* and reflected from the mirror appear to come from *H′*.

Consider a ray from *H*. In accord with the laws of reflection, this ray, the perpendicular to the mirror's surface, and the reflected ray are all in the same plane. Figure 4 – 4 is drawn in this plane and in it the angles of incidence *i* and reflection *r* are equal. The plane itself is perpendicular to the surface of the mirror. Consequently the perpendicular to the mirror from *H* is also included in the same plane. This perpendicular is the line *HM* in the figure. The dotted extension of the reflected ray cuts this perpendicular at some point *H′* (still in the same plane).

Now the angles marked *1* and *2* in the figure are equal because angles *1* and *i* are equal and angles *r* and *2* are equal. Therefore, the two right triangles have all their corresponding angles equal and they have a common side *LM* between corresponding angles. Hence, the corresponding sides *H′M* and *HM* are equal. *H′* is therefore as far behind the mirror (on the perpendicular through *H*) as *H* is in front of the mirror.

We can repeat this reasoning, using another ray leaving *H* in a different direction. The point *L* on the mirror will be different, but the points *H* and *M* are the same. Consequently the right triangle *HML* will still be congruent to the right triangle *H′ML*. Therefore, the point *H′* will be the same as before. We can apply the argument to a third ray and so on. Therefore all the rays which originated from *H* and were reflected by the mirror appear to diverge from *H′*. The point *H′* is the image of *H*. This image *H′* is on the perpendicular to the mirror through *H* and is the same distance behind the mirror as *H* is in front of the mirror.



4–4 The geometry of a single ray reflected from a plane mirror.

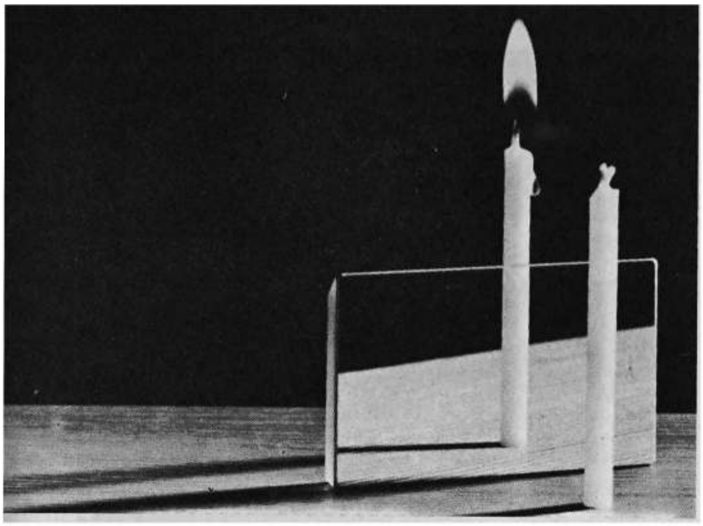
Summing up our findings, we have shown that those rays which start originally at the point *H* and are reflected by the mirror appear to diverge from the point *H′* located on the extended perpendicular from *H* to the mirror. The point *H′* is as far behind the mirror as *H* is in front of it. As far as the eye is concerned, then, there appears to be a point source of light at *H′*. We speak of this point as the image of *H*.

By the same reasoning we can locate the image of any point on an arrow, and thus find the complete image of the arrow. Every point on the image is exactly opposite the corresponding point on the object and equally far from the mirror surface. The image and object are exactly similar and of the same size.

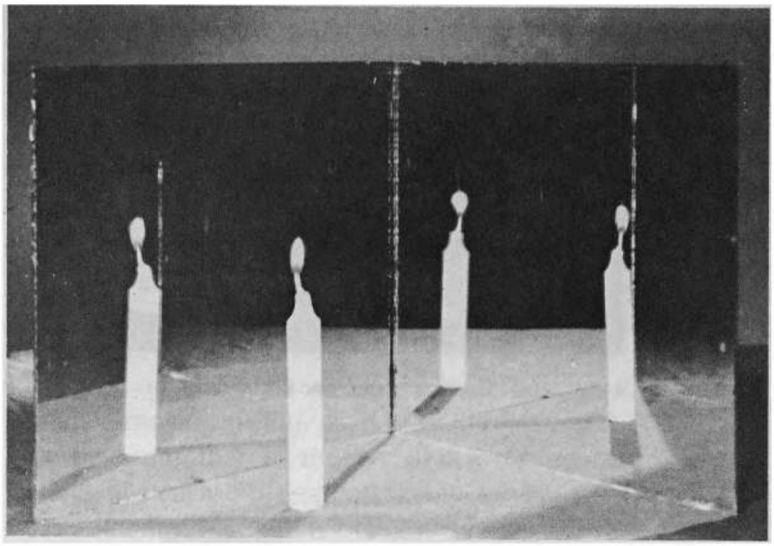
In our construction we have extended the reflected rays behind the mirror, but it is quite clear that we could not find the light from the arrow by going behind the mirror. Because the mirror causes the rays to appear as if they come from the points of the image, although they actually do not, we say that a *virtual image* is formed. We use this term to distinguish it from a *real image,* which we shall meet later.

We have obtained the location of the virtual image by a calculation based on the two laws of reflection. It is easy to verify the results of this calculation by a simple experiment. A small mirror is mounted perpendicular to a tabletop, and an object having a height somewhat greater than the mirror is placed in front of it, as shown in Fig. 4 – 5. An identical object is then placed behind the mirror and is moved about until its top portion (seen over the top of the mirror) appears as a continuation of the image when it and the image are viewed *from any possible position.* The real object behind the mirror then is in the same position as the virtual image. Measurements show that this is the position given by our calculation.

Now that we have found a way to locate images in one mirror, we can easily locate all the images when there are several mirrors as in



4–5 Location of the virtual image. This experiment verifies our calculations based on the laws of reflection.

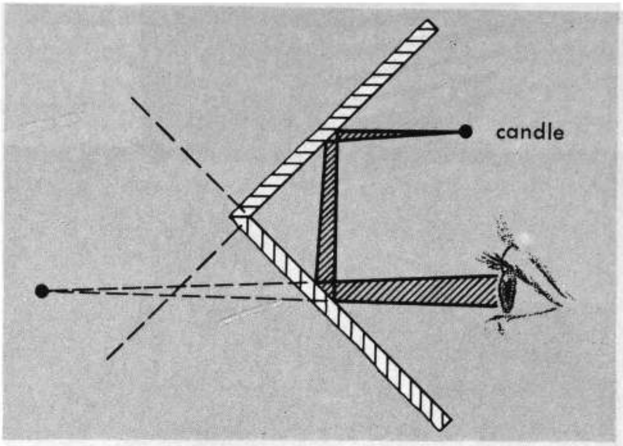


4–6 Multiple reflections in two mirrors. One candle produces three reflected images. Can you explain why?

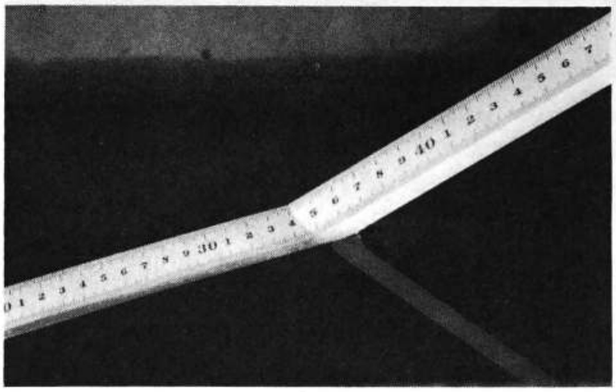
Fig. 4 – 6. For example, in Fig. 4 – 6 we see a candle and its three images formed in the mirrors which are set at right angles to each other. The location of the only image difficult to understand is illustrated in Fig. 4 – 7. By tracing back the "rays" entering the eye, using only the laws of specular reflection, we do find that they diverge from the candle.

4–3 REFRACTION

In order to make an experimental study of reflection we set up apparatus to observe single pencils and beams of light being reflected from surfaces. By systematic observations we arrived at two simple laws of specular reflection. These laws, in turn, make it possible to explain the formation of images in plane mirrors. We thus see that the reduction of a wide range of phenomena to two straightforward laws is a great gain in simplicity and gives us a powerful tool for prediction. It would be difficult to remember the characteristics of all the images we can produce with plane mirrors and a wide variety of positions for the object. If, however, we know how to use these two laws we can arrive at a detailed description of the image formed by any combination of plane mirrors.



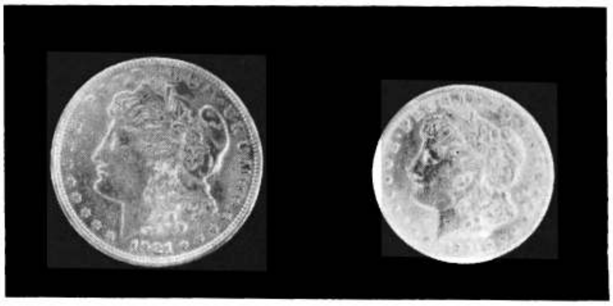
4–7 Formation of an image by double reflection.



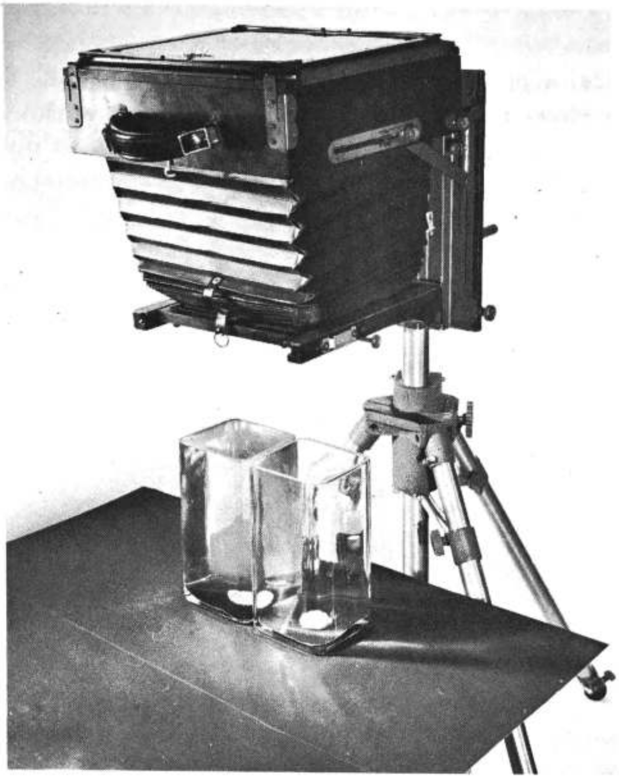
4–8 This photograph shows the apparent bending of a meter stick at the point where it enters water. The gray shape sloping down from the stick is a reflection.

Let us see if a similar study will give us simple explanations of the second group of optical phenomena we observe when light passes into or out of a transparent material. In doing so it often changes its direction. This bending of a light beam as it passes from one material into another we call *refraction*.

Figure 4 – 8 shows a straight stick extending into a tank of clean water. The apparent bending

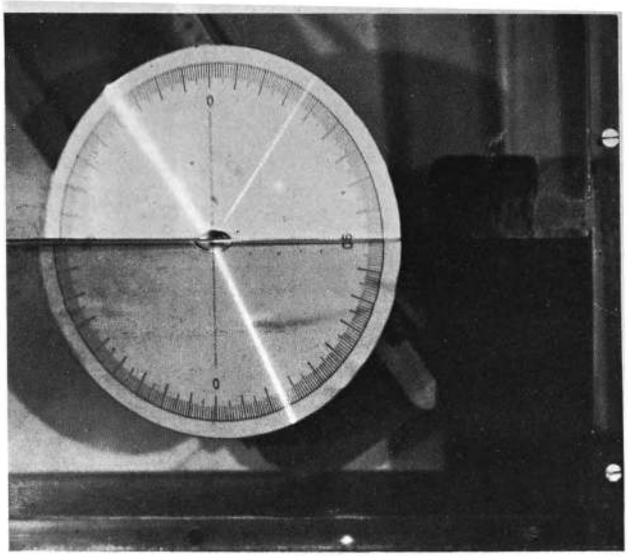


4–9a The "floating" coin. See Fig. 4–9(b).



4–9b The position of camera and containers used in photographing the "floating" coin in Fig. 4–9(a).

of the stick at the point at which it enters the water certainly indicates that something happens to the light which passes from the stick to the camera. Another illustration of the same effect is shown in Fig. 4 – 9 (a). The coin on the right is in an empty container, and an identical coin is in a water-filled container on the left. The coins were photographed simultaneously with the apparatus shown in Fig. 4 – 9 (b). The coin in the water looks nearer to the camera and larger than the other one. This "floating" coin effect and the illusion of the "bent" stick occur because light changes direction when it goes from one material, such as glass or water, into another material, such as air. Although the effect of the bending of light is clearly shown by these observations, they tell us little about what happens to individual pencils of light. We can simplify our study by isolating a single refracted pencil of light as shown in Fig. 4 – 10. This photograph shows a rather broad pencil of light passing through air, and entering the water in an aquarium. Notice that the direction of the pencil changes abruptly at the water surface and some of the light is reflected from the surface, back into the air.



4–10 Light passing from air into water.

We may call the incoming pencil of light the *incident* pencil, as we did in our study of reflection. In Fig. 4 – 10 the angle between this pencil and the normal to the surface is the *angle of incidence.* The light pencil after bending at the water surface is called the *refracted* pencil, and the angle that it makes with the normal is the *angle of refraction.*

By inserting a white plastic protractor into the aquarium, as in the figure, we can learn more about the direction taken by these rays. When we place the protractor perpendicular to the water surface so that the incident rays just glance along the protractor surface, the refracted rays just glance along the protractor surface also. We may repeat this experiment with different angles of incidence and different pairs of materials. For most materials we find the same result. It is summarized in the first law of refraction:

*The incident ray, the refracted ray, and the normal to the surface are all in the same plane.**

Notice the similarity between the first law of refraction and the first law of reflection (Section 4 – 1).

4–4 EXPERIMENTS ON THE ANGLES IN REFRACTION

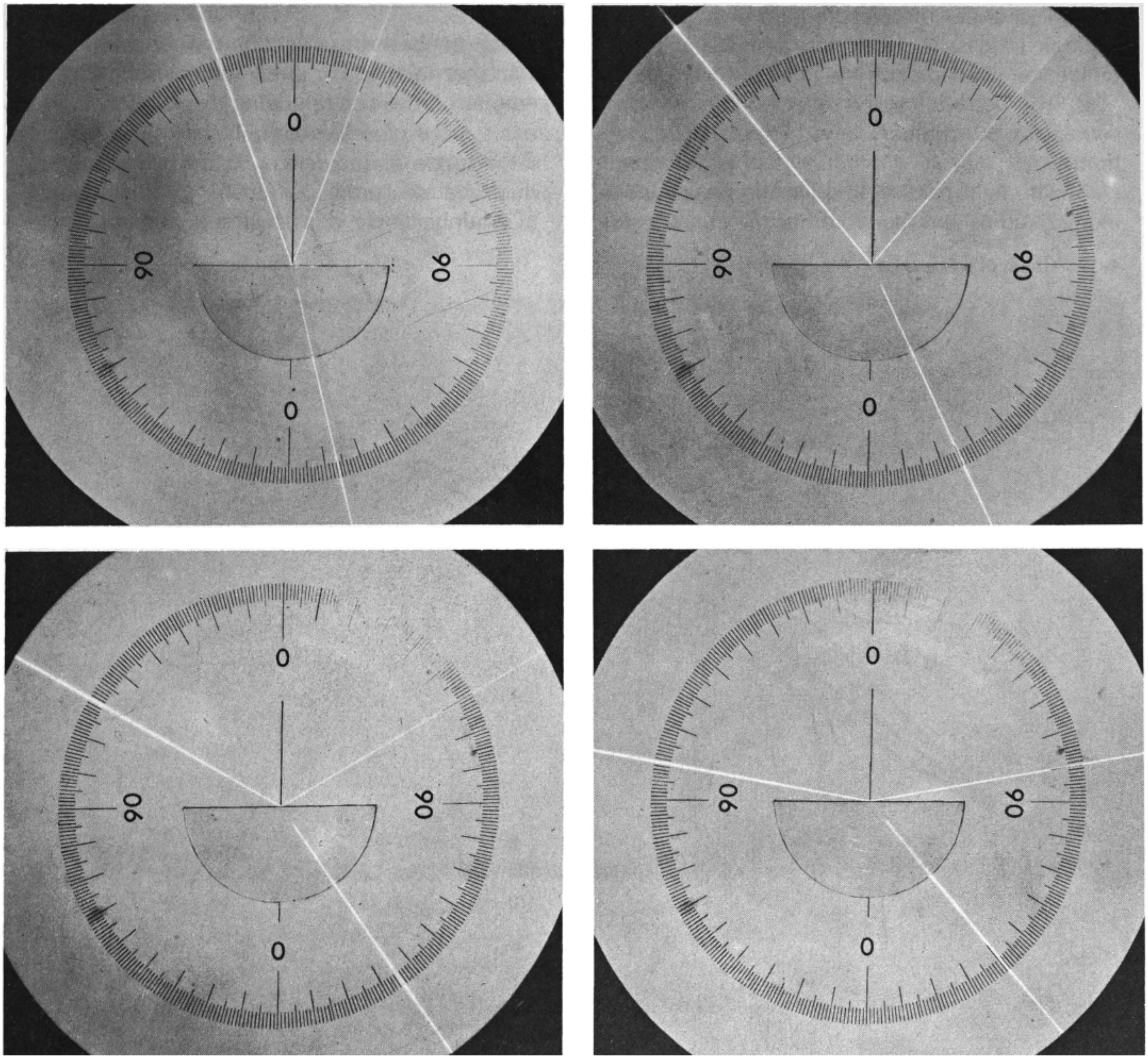
The second law of reflection tells us that the angles of incidence and reflection are equal. Will experimental observation help us to arrive at some similar relationship between the angles of incidence and refraction?

If we wish to make reasonably accurate measurements of the angles of incidence and refraction, we need narrower pencils of light than those shown in Fig. 4 – 10, so that we can approximate single rays. We must change the angle of incidence through a wide range and measure the various angles of refraction. Finally, we must do all this for various pairs of materials to find a general law.

Using air and glass we carried out the experiment indicated in Fig. 4 – 11. We constructed a light source that is moved easily through an angle of nearly 90 degrees and is adjustable so that the narrow pencil of light coming from it can just glance along the surface of a piece of white paper to make the pencil visible. A semicircle of glass, placed in the path of the pencil of light as shown, served as the second material. The back of this plate was ground until slightly

---

*In some crystalline substances the incident ray, the normal to the surface, and the refracted ray are not in the same plane.



4–11 The refraction of light entering glass at different angles of incidence.

rough and then painted white so that the pencil is visible as it traverses the glass. The glass semicircle is mounted on a circular scale with both centers at the same point, and the normal to the diameter of the semicircle passes through zero on the scale indicating the number of degrees. The incoming narrow ray is directed exactly at the common center. We can therefore measure the angle of incidence and the angle of refraction associated with it. The reflected ray is also visible in the pictures, and we can use the measured angle of reflection to check the accuracy of alignment.

Pictures were made with this experimental arrangement with the pencil of light adjusted to angles of incidence of 0°, 10°, 20°, 30°, 40°, 50°, 60°, 70°, and 80°. Some of the pictures are shown in Fig. 4 – 11, and all the data appear in Table 1. Study this series of pictures carefully. Do you agree that the following conclusions can be reached?

(1) The angle of refraction for light passing from air to glass is always less than the angle of incidence, except when the latter is 0°.

(2) The behavior of the reflected pencil is described by the ordinary laws of specular reflection.

(3) The light passing through the glass travels along a radius and therefore hits the curved surface perpendicularly. It is not refracted at this boundary but goes straight out as it passes again into the air.

TABLE 1

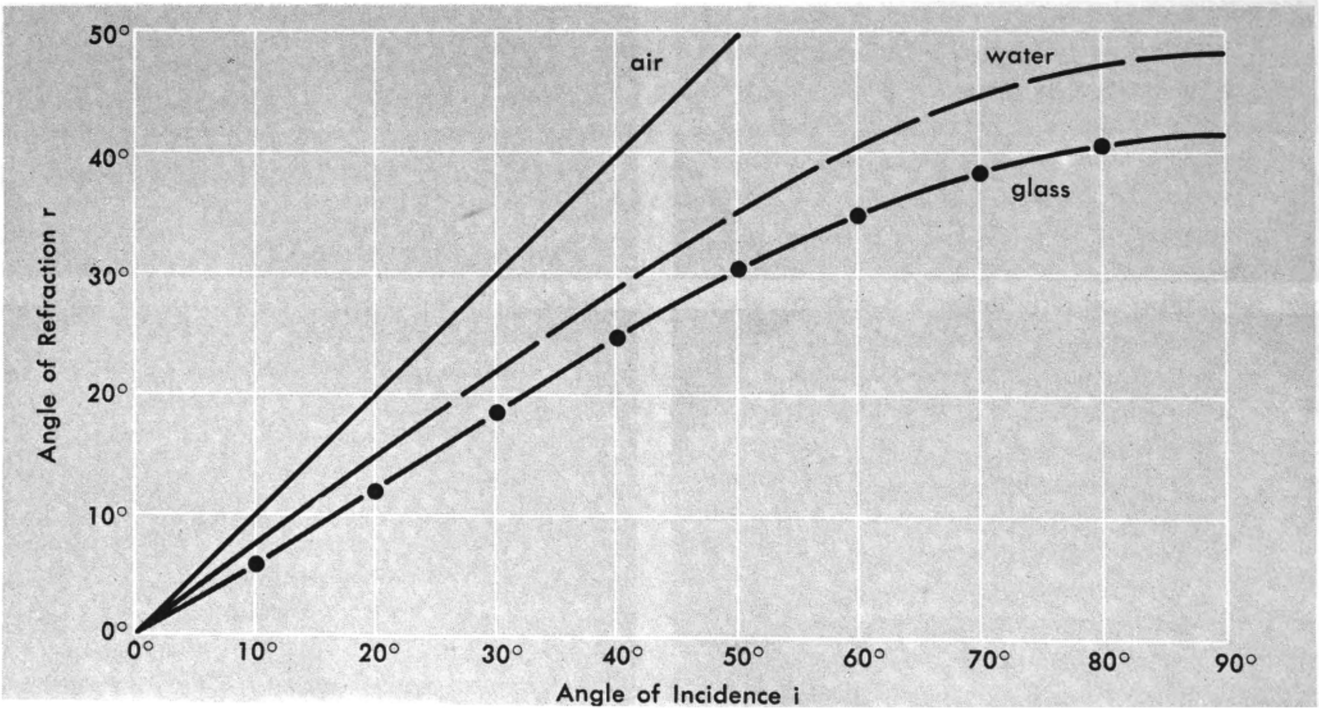
ANGLE OF INCIDENCE, i (DEG)	ANGLE OF REFRACTION, r (DEG)	RATIO, i/r
0	0	Indeterminate
10	6.7	1.50
20	13.3	1.50
30	19.6	1.53
40	25.2	1.59
50	30.7	1.63
60	35.1	1.71
70	38.6	1.81
80	40.6	1.97

Relation between the angle of incidence, i, *and the angle of refraction,* r, *for the passage of light from air to glass*

Each of these observations will be important to us eventually. For the moment let us concentrate on the first, studying the relation of the angles more closely. Table 1 shows the values of the angle of incidence, *i,* and refraction, *r,* which we measured.

Certainly, angle *r* is always smaller than the

4–12 Graph of r versus i.



corresponding angle i. We might try to investigate the ratio i/r to see whether it is constant, but we find, as shown in the last column of Table 1, that the ratios vary from about 1.5 at small angles i to more than 1.9 when $i = 80°$. The ratio is not a constant, but the angles of refraction do vary in a regular way with the angles of incidence. This regular relation becomes clear when we plot the graph shown in Fig. 4 – 12. This is the graph of the angle r versus angle i. It was made by plotting the experimental points, which are marked, and drawing a smooth curve through the points.

We can use the graph to predict what the angle of refraction would be for any given angle of incidence. For example, we can be quite sure what angle of refraction we shall find when the angle of incidence is 45°. From the graph we get 27.8°. A measurement taken with the same apparatus with which we got our original data gave the value 27.6°, in close agreement with our expectation.

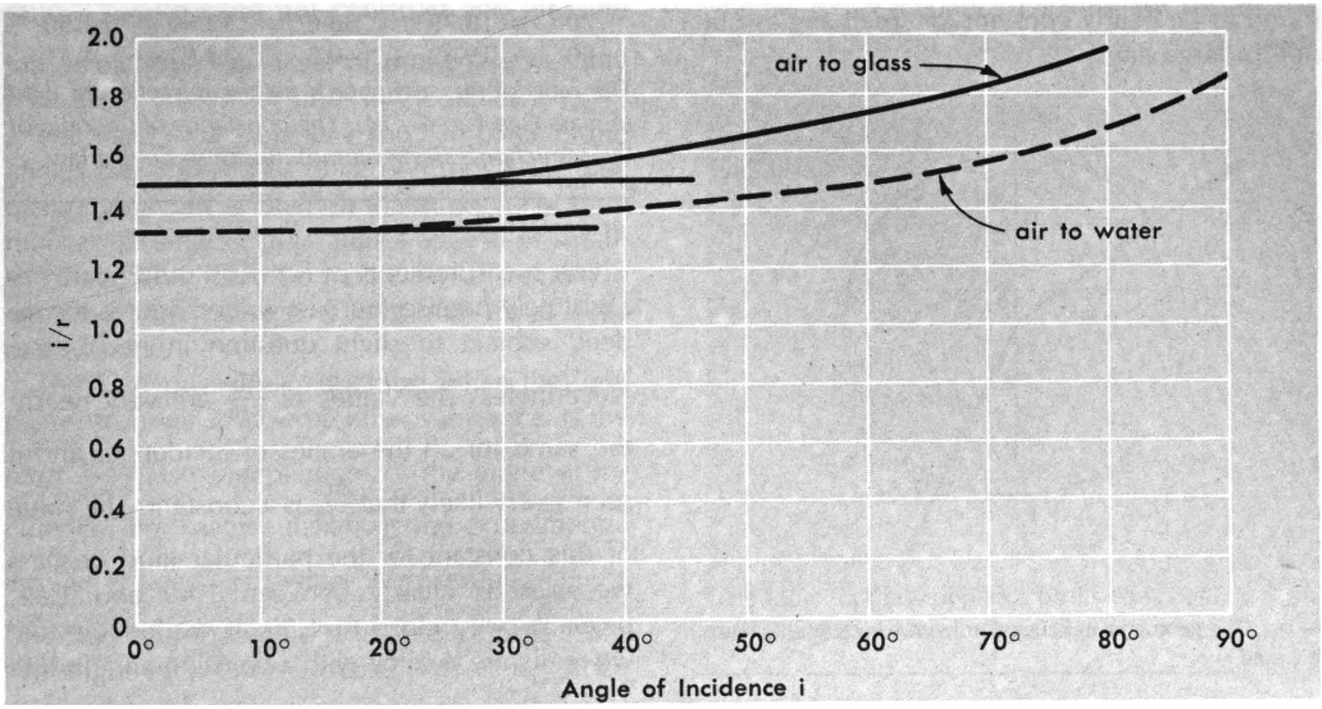
Now a question naturally arises. Is the graph we have drawn correct for all pairs of materials that we might use, or is it true only for the passage of light from air into glass? To answer this question we may try a new experiment with a different pair of materials. The dashed line in Fig. 4 – 12 shows the graph obtained from a series of measurements made with light passing from air into water. Notice that for a given angle of incidence the angle of refraction is always greater in water than it is in glass. Furthermore, when light just passes from air into more air of the same density, the pencil of light is not bent and therefore $r = i$. This gives a straight-line graph at 45° slope instead of a curve that bends over. Evidently, different pairs of materials lead to different graphs of r versus i.

4–5 THE INDEX OF REFRACTION: SNELL's LAW

It is possible to perform a series of similar experiments, using many different materials, and to plot a whole series of graphs of the type shown in Fig. 4 – 12. A book full of these graphs could be very useful in describing just how light is refracted in any pair of substances that have been studied. But it is less convenient than having a single law.

We have already tried the relation $\frac{i}{r}$ = constant and found that it does not hold for all angles. But, as Fig. 4 – 13 shows, for small angles of incidence the law is that $\frac{i}{r}$ is constant, about 1.5 for glass and 1.33 for water. At larger angles the constancy fails, and $\frac{i}{r}$ rises.

4–13 Graph of i/r versus i.

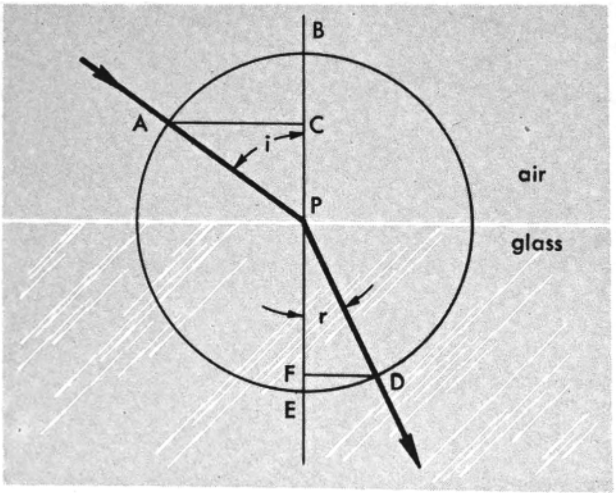


One way to look for a convenient representation of the relation of r to i is to try to extend the constancy of the ratio $\frac{i}{r}$ by finding simple functions of i and r that maintain the constant ratio at all angles. Following this program we would look for functions which are proportional to the angles when they are small and which compensate for the rise in $\frac{i}{r}$ when the angles get large. That we shall find such functions is not clear before we start; in fact, data on refraction were known to Ptolemy and used for a thousand years before an adequate simple law of refraction was stated. Finally in 1621 Willebrord Snell found a beautiful way to describe the relation of i and r. The same relation was published by Descartes in 1638 in the form we now use.

We can explain the discovery of Snell and Descartes with the aid of Fig. 4 – 14. In the figure a ray of light is indicated coming in from the upper left. It enters the glass at the point P where it is refracted. A circle has been drawn in the plane of the rays with its center at P; and the normal to the surface of the glass is shown. The arcs AB and ED' on the circumference of the circle are proportional to the angles i and r. Therefore

$$\frac{i}{r} = \frac{AB}{ED}$$

This is the ratio plotted in Fig. 4 – 13, the ratio found to be nearly constant for small angles but not for large ones.



4–14 The geometry of refraction from which Snell's law is found.

In effect, what Descartes and Snell did was to try to relate i and r by examining the ratio of the semichords AC and FD. That is, in place of $\frac{AB}{ED}$ they calculated $\frac{AC}{FD}$ for a wide range of angles. For small angles, their ratio is the same as $\frac{i}{r}$ because the chords and arcs are almost equal, but for large angles the chords and arcs are substantially different so that the chord ratio behaves differently. Let us look at this chord ratio for our experimental data taken with glass (Table 1).

We can find $\frac{AC}{FD}$ by measuring the semichords in a series of figures like Fig. 4 – 14. We make the figures with the measured angles i and r from our experimental results. Measuring the semichords and calculating their ratio gives the results shown in Table 2.

TABLE 2

i *(DEG)*	AC *(CM)*	r *(DEG)*	FD *(CM)*	$\frac{AC}{FD}$
10	3.47	6.7	2.30	1.50
20	6.84	13.3	4.60	1.49
30	10.00	19.6	6.71	1.49
40	12.83	25.2	8.52	1.51
50	15.32	30.7	10.20	1.50
60	17.32	35.1	11.50	1.51
70	18.79	38.6	12.50	1.50
80	19.69	40.6	13.02	1.51

The ratio of the semichords for the refraction of light in glass (see Fig. 4—14). The constancy of the ratio of the semichords (last column) illustrates Snell's law.

Our measurements of i, r, AC, and FD are not absolutely precise; our last written figures are indeed subject to slight question in each case. Nevertheless the values of $\frac{AC}{FD}$ are very nearly the same for all the angles of incidence; and it seems very likely that $\frac{AC}{FD}$ is a constant. The value of this constant for the particular kind of glass we used is clearly between 1.49 and 1.51, probably very close to 1.50. If we assume that we really are dealing with a constant and, in line with our data, try the value 1.50, we find that all our measurements are compatible. All agree with 1.50 to within experimental accuracy. Better experiments with smaller experimental uncertainties confirm that we have a constant whose value is close to 1.50.

We can also get the same ratio by using the sines of the angles i and r. In Fig. 4 – 14 we see that

$$\sin i = \frac{AC}{AP} \quad \text{and} \quad \sin r = \frac{FD}{PD}$$

Because AP and PD are both radii of the same circle we can express $\sin r$ as $\frac{FD}{AP}$; and consequently

$$\frac{\sin i}{\sin r} = \frac{AC/AP}{FD/AP} = \frac{AC}{FD}$$

The ratio of the sines of the angles of incidence and refraction is the same as the ratio of the semichords. It is equal to the same constant that we found experimentally for their ratio. In going from air to glass, therefore,

$$\frac{\sin i}{\sin r} = 1.50$$

according to our experiments. Because the sines of all angles have been tabulated, this form is often more convenient to use than the ratio $\frac{AC}{FD}$. We need not construct and measure the semichords on a set of drawings; we only look up the values of sines in the tables.

The use of the sines of i and r in place of the angles has allowed us to carry through our idea of trying to find a function that gives a constant ratio at all angles, large angles as well as small. We have thus found a simple law of refraction describing the passage of light from air to glass at any angle.

The next question is obvious: when the light passes from air into some other substance, is the ratio $\frac{\sin i}{\sin r}$ also independent of the angle of incidence? For many substances the experimental answer is "yes." For water, for example, $\frac{\sin i}{\sin r}$ is constant, and the value of the constant is 1.33. We might even have expected this value, since we already know that $\frac{i}{r}$ for water is about 1.33 for small angles where the ratio of the sines and the ratio of the angles is nearly the same. (See Fig. 4 – 13.) For other substances the constant ratio $\frac{\sin i}{\sin r}$ has different values. The value for a particular substance is a property of the substance like its boiling or melting temperature. It may even help to identify the substance. The constant is called the *index of refraction* of the substance or, to be more exact, the index of refraction for light going from air into the substance. Some of these indices are listed in Table 3.

The law that $\frac{\sin i}{\sin r}$ is constant for all angles of incidence is Snell's law. It tremendously simplifies our description of refraction. If we wish to know how light will be bent on entering any material from air, we need have available only the index of refraction of the material. A collection of these indices, sufficient to include nearly all important substances, can be placed on a single page of a book and can replace many books full of graphs of the type shown in Fig. 4 – 12. Further, when we start on the development of a theory, or model, of light in Chapter 5, we will have a simple and general law of behavior of light that our theory must explain. We can test any model that we invent by seeing whether it can account for Snell's law, the second law of refraction for which we have been searching.

TABLE 3

SUBSTANCE	INDEX OF REFRACTION
Glass*	1.5 – 1.9
Diamond	2.42
Fused quartz	1.46
Quartz crystal	1.54
Glycerin	1.47
Carbon disulfide	1.63
Oleic acid	1.46
Water	1.33

*The index of glass depends on its composition. Most ordinary glasses have indices slightly above 1.5.

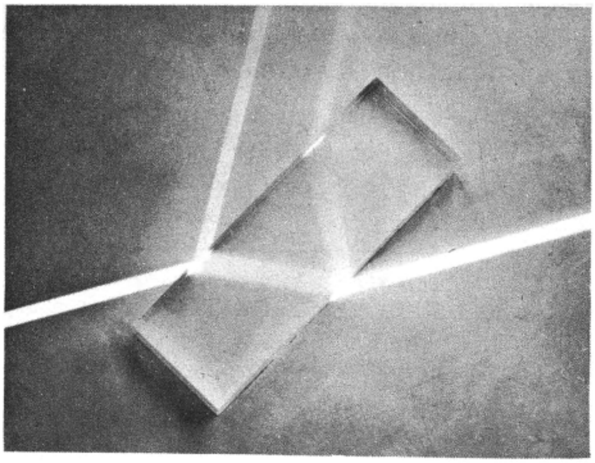
4-6 THE ABSOLUTE INDEX OF REFRACTION

We should note that all our photographs of the refraction of light have been for cases of light passing from *air* into glass, or water, etc. What would we find for the index of refraction when the light travels from a *vacuum* into glass? As a matter of fact, with the limited accuracy of our experimental equipment, we would notice no difference; the same value for the index would result. This we might expect because of the very low density of air. But measurements made with great precision actually show a slightly greater index for light passing from a vacuum into glass than for air to glass. For a sample of glass with index 1.50000 in air, the index in a vacuum is 1.50044.

The index of refraction of any substance in a vacuum is called its *absolute* index of refraction. The value of the index from air to the substance is so close to the value of the absolute index that we rarely need to distinguish between them.

4-7 THE PASSAGE OF LIGHT FROM GLASS (OR WATER) TO AIR: REVERSIBILITY

Thus far we have concentrated our attention on what happens to light as it enters glass or some other substance from air. Suppose that we now investigate what happens when light goes in the opposite direction, from glass into air. We have already found out that when light reaches the surface of the glass along the normal, the part that goes out enters the air without being bent. To find out what happens when the light approaches the glass-air surface at other angles we may make use of a glass block with parallel surfaces. As we see in Fig. 4-15, a pencil of light that is incident on one side of the glass leaves the glass on the other side traveling parallel to the incident direction. The refraction in going from glass to air, therefore, is exactly the opposite of the refraction in going from air to glass.

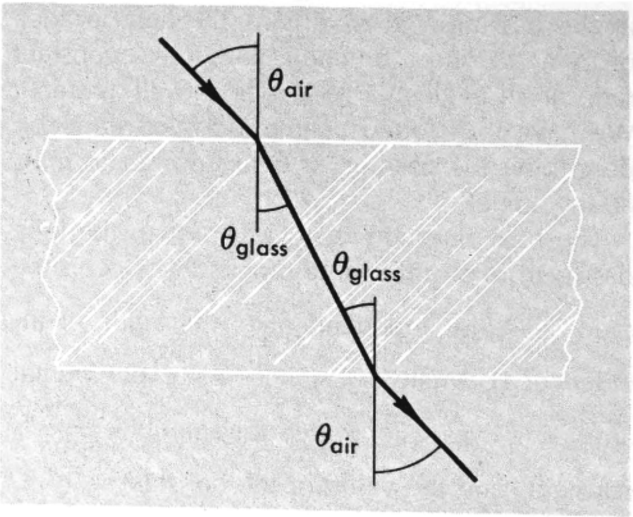


4-15 Light passing through a block of glass with parallel faces.

This situation is illustrated in the ray diagram of Fig. 4-16. Here the light is traveling in the direction shown by the arrows. As it passes from air to glass, the angle of incidence is θ_{air} and the angle of refraction is θ_{glass}. The two are related according to Snell's law by the equation

$$\frac{\sin \theta_{air}}{\sin \theta_{glass}} = n_{glass}$$

where n_{glass} is the index of refraction of light passing from air into glass (very nearly the absolute index of glass). When the light approaches the second surface on its way from glass to air, it is incident on the surface at the angle θ_{glass} because the two surfaces are parallel. Furthermore we found experimentally



4-16 The angles and their relations for light passing through a block of glass with parallel faces.

that the rays entering and leaving the glass are parallel; and since the two normals are parallel, the two angles in air are equal. Therefore the rays leave the glass at the angle θ_{air}. We can therefore conclude that in leaving the glass the angles of incidence and of refraction are related as

$$\frac{\sin i}{\sin r} = \frac{\sin \theta_{\text{glass}}}{\sin \theta_{\text{air}}} = \frac{1}{n_{\text{glass}}}$$

This is Snell's law again, the ratio of sin *i* to sin *r* is constant; and the index of refraction for light going from glass to air is just the inverse of the index for light going the other way. Because one index is the inverse of the other, we do not need two indices of refraction, one from air to glass and another from glass to air. We can use the single air-to-glass index and write the relation of the angles in passing either from air to glass or from glass to air as

$$\sin \theta_{\text{air}} = n_{\text{glass}} \sin \theta_{\text{glass}}$$

From experiments in which light goes from other substances into air we conclude that Snell's law applies equally well to light entering or leaving a substance. It is therefore convenient to write the law in the general form

$$\sin \theta_{\text{air}} = n_{\text{m}} \sin \theta_{\text{m}}$$

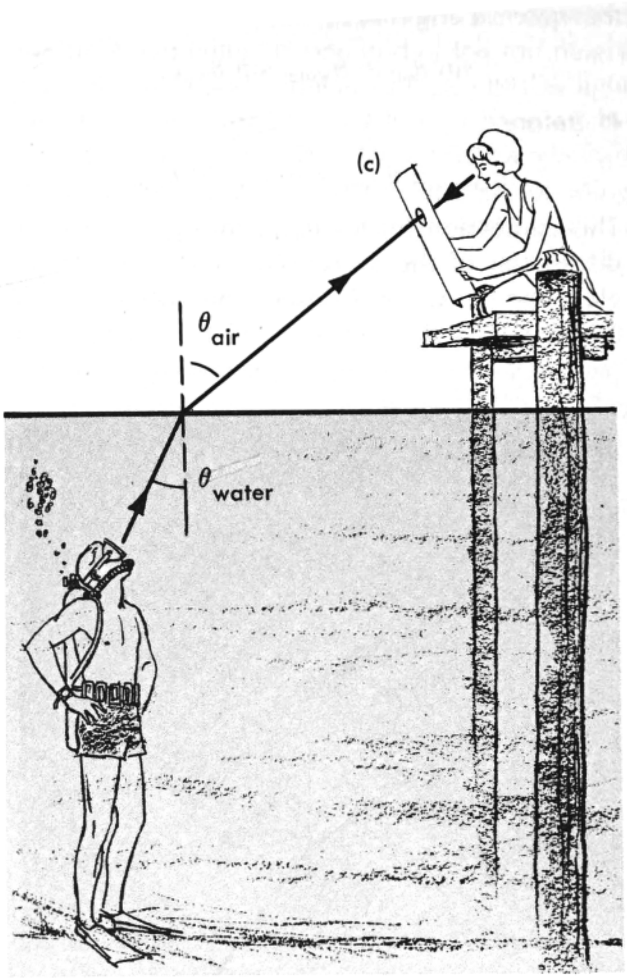
where n_{m} is the index for light going from air into the material *m*, and θ_{m} is the angle between the normal to the surface and the rays in that material. The letter *m* stands for the name of the material. For example, for light traveling from air to diamond or diamond to air, we will have

$$\begin{aligned} \sin \theta_{\text{air}} &= n_{\text{diamond}} \sin \theta_{\text{diamond}} \\ &= 2.42 \sin \theta_{\text{diamond}} \end{aligned}$$

The general relation, $\sin \theta_{\text{air}} = n_{\text{m}} \sin \theta_{\text{m}}$, implies that light paths described by Snell's law are reversible. In this respect they are just like light paths in which specular reflection takes place. You can easily convince yourself of this reversibility with experiments like this: Suppose one of us is under water and the other is above. As in Fig. 4 – 17, we are looking each other in the eye. From the general form of Snell's law, the path of the rays of light going through the water's surface is described by

$$\sin \theta_{\text{a}} = 1.33 \sin \theta_{\text{w}}$$

regardless of whether θ_{a} or θ_{w} is the angle of incidence; thus the light rays from your eye to mine should follow the same path as those from my eye to yours. Now comes the test. We arrange a barrier (*c* in Fig. 4 – 17) with a small hole in it so that you can just see my eye through it. I then find that this is just the position of the barrier that also allows me to see your eye. Many experiments with many different substances confirm the prediction of Snell's law and the laws of reflection that such light paths are reversible.



4–17 The reversibility of light paths.

4–8 THE PASSAGE OF LIGHT FROM GLASS TO WATER

We can now predict what happens when light goes from glass to water. Suppose for a moment there is a layer of air between two parallel surfaces of glass and water as in Fig. 4 – 18. Then as the light leaves the glass, it goes into the air at angle θ_{air} such that

$$\sin \theta_{\text{air}} = n_{\text{glass}} \sin \theta_{\text{glass}}$$

It next enters the water and proceeds through the water at angle θ_{water} such that

$$\sin \theta_{\text{air}} = n_{\text{water}} \sin \theta_{\text{water}}$$

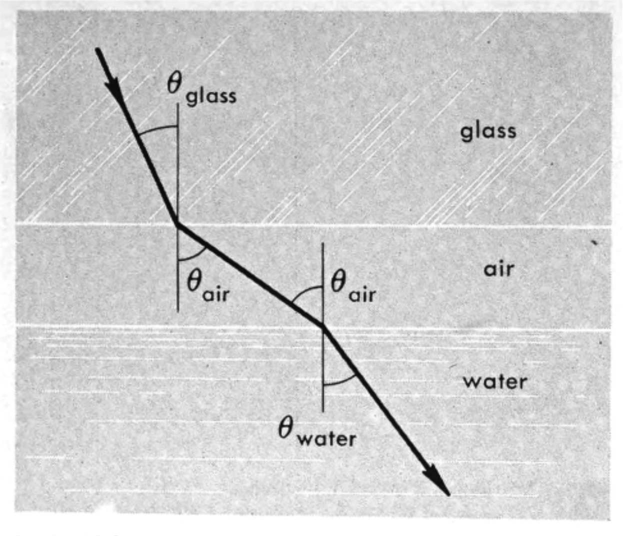
Therefore

$$n_{\text{glass}} \sin \theta_{\text{glass}} = n_{\text{water}} \sin \theta_{\text{water}}$$

This expression relates θ_{water} to θ_{glass}, and is independent of the thickness of air between the glass and the water. Consequently we may hope that it relates θ_{glass} to θ_{water} in the special case when the thickness of the air layer is zero; that is, when the water and glass are touching. Indeed, if you try it, you will find that this relation holds. Further, it is not restricted to glass and water. In general, for two materials, 1 and 2, for which Snell's law holds

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

no matter which way the light is going. The symmetry of this form of Snell's law shows the reversibility of light paths as they pass any surface for which the law is valid.



4–18 Light passing from glass to water through a layer of air.

It may seem that in going from the form $\frac{\sin i}{\sin r} = \textit{const}$ to $n_1 \sin \theta_1 = n_2 \sin \theta_2$, we have substantially modified Snell's law. That this is not the case, we can see by a slight rearrangement of the above equation. In going from material 1 to material 2, θ_1 is the angle of incidence, and θ_2 is the angle of refraction. Our general statement can then be read

$$\frac{\sin i}{\sin r} = \frac{\sin \theta_1}{\sin \theta_2} = \frac{n_2}{n_1}$$

It still says that the ratio of the sines of the angle of incidence and refraction is a constant since $\frac{n_2}{n_1}$ is constant. This constant is called the *relative index* of the two materials or the index of refraction of light passing from material 1 into material 2. It is often written n_{12} (read $n_{\text{one two}}$).

The statement

$$\frac{\sin \theta_1}{\sin \theta_2} = n_{12}$$

therefore, takes us back to the basic statement of the law.

Moreover, since

$$n_{12} = \frac{n_2}{n_1}$$

we can find the relative indices for any pairs of materials in terms of their absolute indices. From the relatively few absolute indices, all the relative indices can be computed. For example, the relative index for light going from water to glass is

$$n_{\text{wg}} = \frac{n_{\text{g}}}{n_{\text{w}}} = \frac{1.50}{1.33} = 1.13$$

4–9 TOTAL INTERNAL REFLECTION

If we apply Snell's law blindly to all possible cases, we soon get into trouble. A difficulty arises when we try to compute the direction of the refracted ray for light traveling from a medium with a higher absolute index of refraction to a medium with a lower index. For example, consider light passing from glass, with $n = 1.50$, to air, at the particular angle of incidence of 41.8°. The sine of this angle is 0.667. The sine of the angle of refraction is given by Snell's law as

$$\sin \theta_{\mathrm{air}} = n_{\mathrm{glass}} \sin \theta_{\mathrm{glass}}$$

Hence

$$\sin \theta_{\mathrm{air}} = 1.50(0.667) = 1.00$$

When $\sin \theta = 1$, θ is 90°; therefore θ_{air} is 90°. In other words, the refracted light emerges from the glass in a direction parallel to the surface; it just glances along the surface. This angle of refraction is the largest one possible. What if a beam strikes the surface at an angle of incidence greater than 41.8°?

Let us try to set up an experiment that will tell us what actually happens. We can do this with the arrangement shown in Fig. 4 – 19, in which several pencils of light, diverging from one another, enter the left side of a glass prism. After refraction at the first surface, they meet the surface at the top of the prism at different angles. Notice that the pencils in the right-hand part of the beam do not result in any refracted pencils; instead, they are totally reflected back into the glass. This phenomenon is known as *total internal reflection*. The smallest angle of incidence for which total internal reflection occurs is called the *critical angle*. If you measure the angle of incidence for the last two pencils on the right you will find that they are larger than 41.8°.

In general for light going from medium 1 into medium 2 as in Fig. 4 – 19, the critical angle in medium 1 occurs when $\theta_2 = 90°$; then from Snell's law,

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

we get

$$n_1 \sin \theta_c = n_2$$

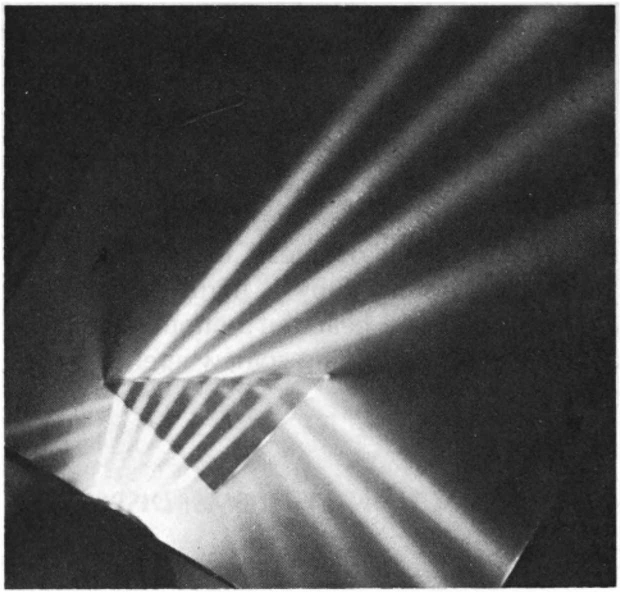
or

$$\sin \theta_c = \frac{n_2}{n_1}$$

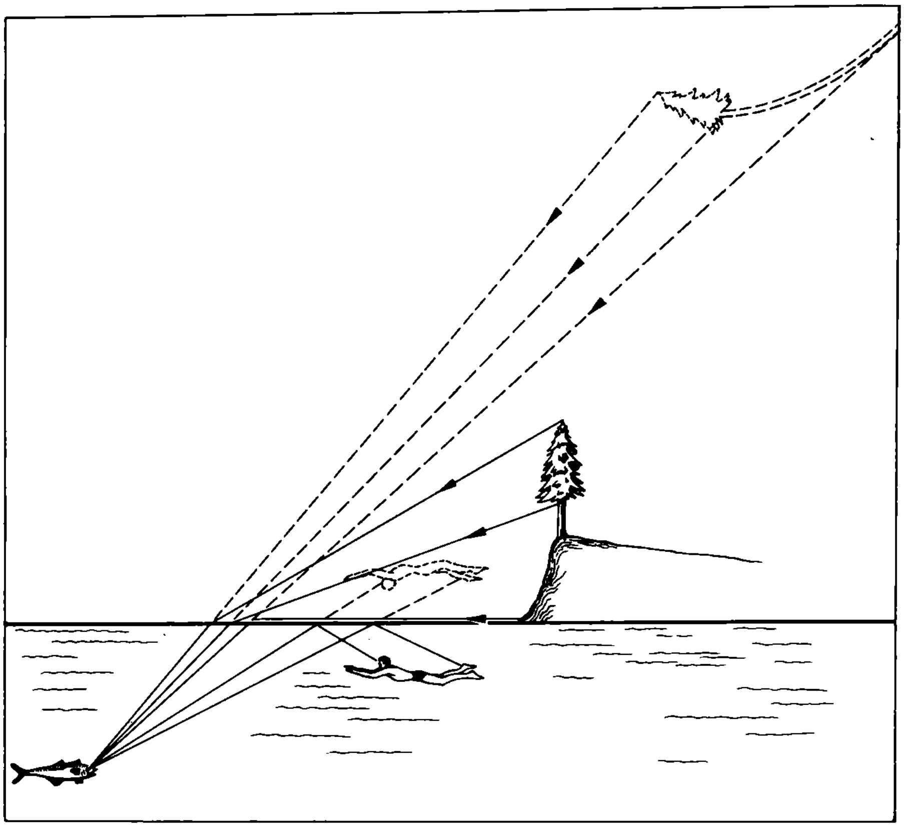
Clearly, total reflection can occur only when n_1 is greater than n_2.

It may seem strange that light stops being refracted and is totally reflected when the critical angle is reached. In fact there is no sudden change. As you see in Fig. 4 – 19, when the angle of incidence at the glass-to-air surface increases, more light (a brighter beam) is reflected and less light (a dimmer beam) is refracted; and as the critical angle is approached, we arrive at the point where all the light is reflected and none transmitted.

Sometimes light is piped around corners by means of total internal reflections. The "light pipe" consists of a cylindrical rod of transparent plastic. Light enters at one end of the rod nearly normal to the end surface. Any part of this light which reaches the side walls will have an incident angle greater than the critical angle and therefore will not escape into the surrounding air. Instead, a succession of total reflections will carry it around in the rod and it will finally emerge at the far end.



4–19 Pencils of light entering a glass prism from the lower left. The two pencils entering nearest the bottom of the prism are totally reflected at the horizontal glass-to-air surface; the other four pencils are partially reflected and partially refracted at this surface.



4–20 The effect of refraction and total reflection on images seen by a fish. The rays of light from the tree appear to be coming from the virtual image up in the sky. Because this image is far away, the fish may not notice that the trunk stretches off toward an infinite distance. From his point of view it is foreshortened. The fish can hardly see any of the ground, because it is compressed into such a small angle in the image. (The images of the swimmer and of the tree were drawn by following more than one ray from each of several points, such as the top of the tree, its base, and so on. Only one of the rays is shown from each point.)

An interesting example of total internal reflection may be seen by a fish that looks upward through the surface of a large, smooth pond or lake (Fig. 4 - 20). He sees the trees around the pond hanging overhead in a small circle. They are surrounded by a mirror that reflects a man swimming near the bottom so that he appears to be swimming somewhere above the surface.

4–10 REFRACTION BY PRISMS; DISPERSION

When we send light through a piece of glass having parallel faces, the rays emerging are parallel to the incident rays. We can, however, change the direction of a light beam by using a piece of glass with two nonparallel faces (Fig. 4 - 21). In fact, whenever the two faces are not parallel, the light will emerge in a new direction.

Furthermore, a careful examination shows that, even when the incident beam is made of parallel pencils of light, the beam emerging from the prism diverges, or spreads out (Fig. 4 - 22). To investigate this spreading, which does not seem to be quite consistent with the laws of refraction, we shall allow the light to travel a considerable distance from the prism and then examine it. We shall use a very narrow incident pencil so that the spreading will be large compared to the pencil width. A simple arrangement for doing this is shown in Fig. 4 - 23 (a). Light from a distant white source, such as the sun or an incandescent light bulb, is allowed to pass through a narrow aperture in an opaque barrier. This aperture produces a narrow beam of light. A screen is set up at some distance from the prism. It is found that the light falling on the screen is no longer "white." Instead, a brilliant spectrum of colors —violet, blue, green, yellow, orange, and red— is spread across the screen. The colors are like those that we see in a rainbow, with red at one end and violet at the other. The red deviates least from its original direction, and the violet most.

The amount of deviation produced by a prism is determined by the angle between the surfaces through which the light passes, by the direction of incidence on the first face, and by the index of refraction of the prism. Of these, the only quantity that can differ for the different colors of light is the index of refraction. We are thus forced to conclude that we must give up some of the simplicity achieved in our earlier experiments. Having decided that we could describe the refractive properties of any substance in terms of a single number, its refractive index, we now find that the refractive index appears to depend on the color of the light! Apparently white light is made up of light of different colors; and a material such as glass has different refractive indices for the different colors.

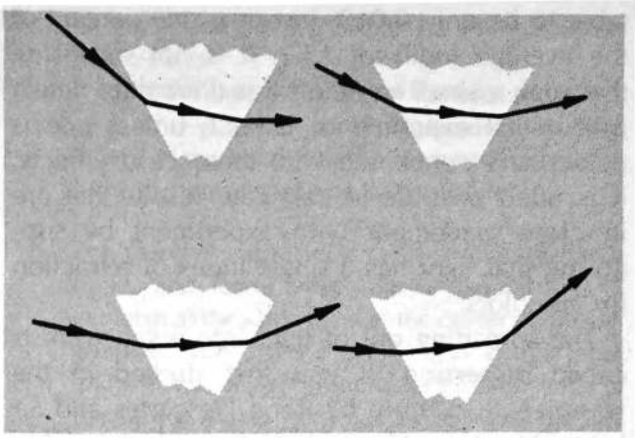
If this explanation is correct, then we should expect to find the spreading of white light into colors in some of our previous experiments on refraction. Indeed, with careful examination of the refracted beam, it is possible to see a slight separation of colors in both the experimental arrangement of Fig. 4 – 11 and also that of Fig. 4 – 15. In the first case, it is necessary to examine the light at a considerable distance from the semicircle of glass; in the second case it is necessary to use a very thick glass block; in both cases the effect is greatest for large angles of incidence. But the important thing is that the effect is present. It does indeed seem to be related to the nature of the refractive index, not to some special property of a prism. The prism is simply a convenient shape for amplifying the effect into something that is easily visible.

When a beam of white light strikes the prism as illustrated in Fig. 4 – 23 (b) the violet light is deviated most, while the red light is deviated least. Now, no matter which way light passes between air and another medium, Snell's law tells us that the bending of the light is greater the greater the index of refraction of the medium. Consequently the index of refraction of the glass is greater for violet light than it is for red. The variation with color of the index of refraction for a glass that is used in many lenses is given in Table 4.

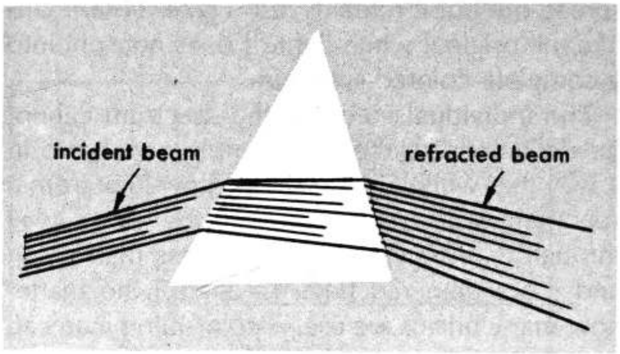
TABLE 4 *Index of refraction of "crown" glass*

COLOR:	VIOLET	BLUE	GREEN	YELLOW	ORANGE	RED
Index:	1.532	1.528	1.519	1.517	1.514	1.513

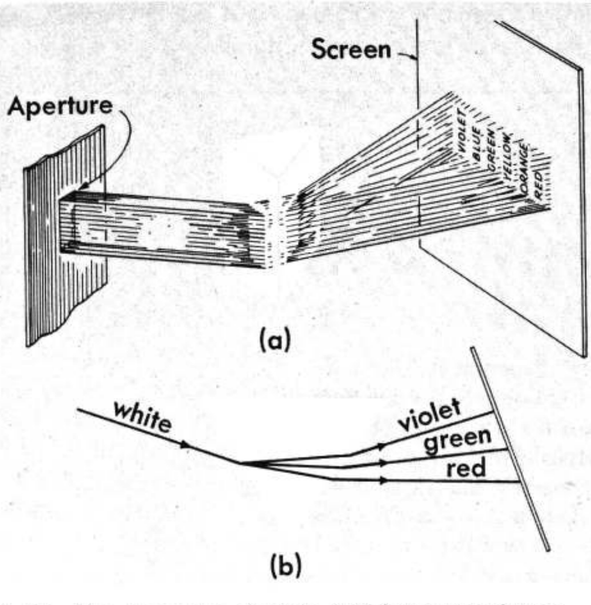
The difference between the index of refraction for violet and for red light is seen from the



4–21 Light bent by passing through a glass block with nonparallel faces.



4–22 A beam of white light diverging as a result of passing through a prism. The divergence is exaggerated.



4–23 The dispersion of white light by a prism into a colored spectrum.

table to be only 0.019, just over one percent of the average index of 1.52. It is not surprising that such a small variation was difficult to detect in our first experiments. In fact, unless one is particularly concerned with these color effects, it is often possible to calculate results that are in close agreement with experiment by supposing that light has a single index of refraction for all colors.

The spreading out of light into a spectrum is called *dispersion*. It was first studied in the seventeenth century by René Descartes and Sir Isaac Newton. Newton performed the additional experiment of trying to break up one portion of the spectrum by inserting a prism in a beam of light of a particular color, say red. All that happens is that the red is slightly further spread out but it remains red [Fig. 4-24 (a)]. Unlike the original white light, it does not split into a complete colored spectrum.

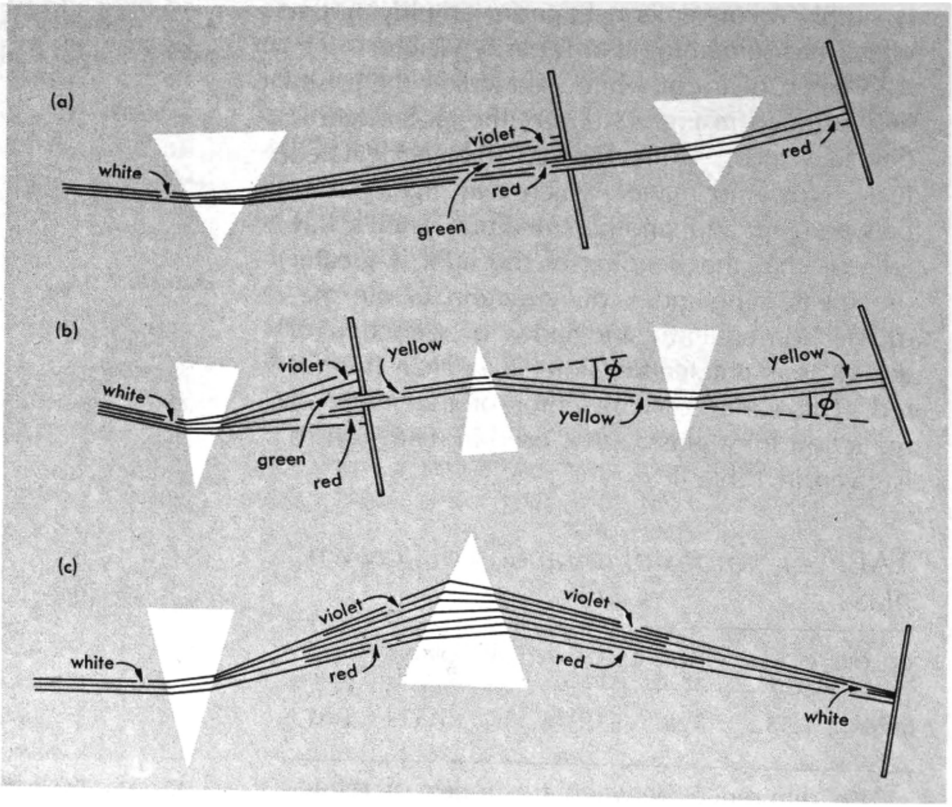
The individual colors of the spectrum cannot be decomposed; they are simple and basic in a way that white light is not. Yellow light from a spectrum will always deviate the same way through a prism. It will deviate less than green and more than red [Fig. 4 – 24 (b)], no matter how many prisms we use. A color-blind man can use the deviation to distinguish colors of light.

Our conclusions are supported by a further experiment of Newton's proving that white light is a combination of many colors. He recombined the spectrum to make white light. We can do the experiment by breaking up a narrow beam into a spectrum by a prism, and then placing in the spectrum a second prism with a greater angle between its faces [Fig. 4 – 24 (c)]. Because of its greater angle the second prism deviates the different colors more than the original prism and the light converges again. The pencils of light of different colors then overlap in some region beyond the second prism, and in this region the light on a screen again appears white. This is what Newton found, as you can too if you repeat his experiment with two glass or plastic prisms.

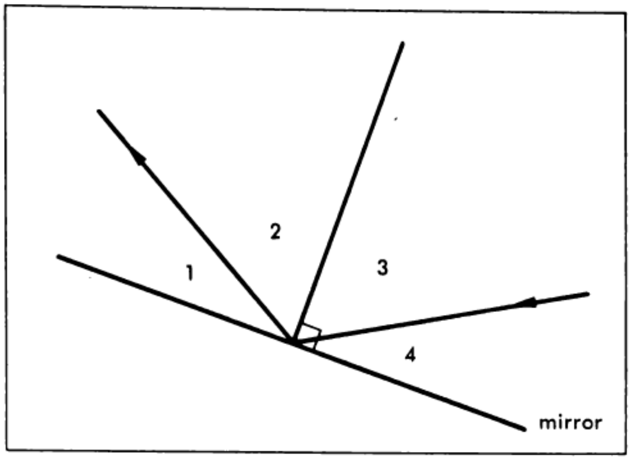
QUESTIONS

1.* In Fig. 4 – 25, which is the angle of incidence and which is the angle of reflection?

2.* How many different rays can strike a mirror at the same point and with the same angle of incidence?



4–24 Experiments with a prism. (a) Light of one spectral color remains the same color on further analysis with a prism. (b) Light of one spectral color is always deviated through the same angle when passing on similar paths through identical prisms. (c) The spectral colors combine to make white light.



4–25 For Question 1.

3.* For a given incident ray of light striking a mirror, how many possible reflected rays are there?

4. If you can see the eyes of someone in a complicated system of mirrors, is it possible for him to see your eyes?

5. A pencil of light can be used as an optical lever to amplify very small motions, as shown in Fig. 1-14, where the expansion of a rod when heated swings a small mirror about a pivot.

(a) If the mirror twists through an angle of .06°, through what angle does the reflected pencil move? What fraction of 360° does this angle represent?

(b) If the distance from mirror to scale is 2m, about how far does the reflected pencil move across the scale?

(c) What effect does the position of the light source have on the amplification?

6. Place a plane mirror (one about 30-40 cm high is convenient) with its center approximately at eye level. Hold a meter stick vertically just in front of your face, with the middle of the stick at eye level; stand in front of and facing the mirror.

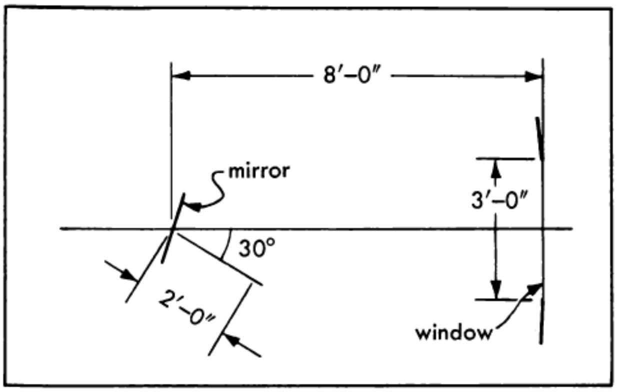
(a) Move toward and away from the mirror. Does your motion change the amount of the stick that you can see?

(b) Formulate a general rule connecting the length of the stick that can be seen with the height of the mirror. By making a ray diagram, show that this rule holds for all distances from the mirror.

(c) Clothing stores often have mirrors that extend all the way to the floor, designed to allow a customer to see his or her full length. Is it necessary for this purpose for the mirror to be as long as it is?

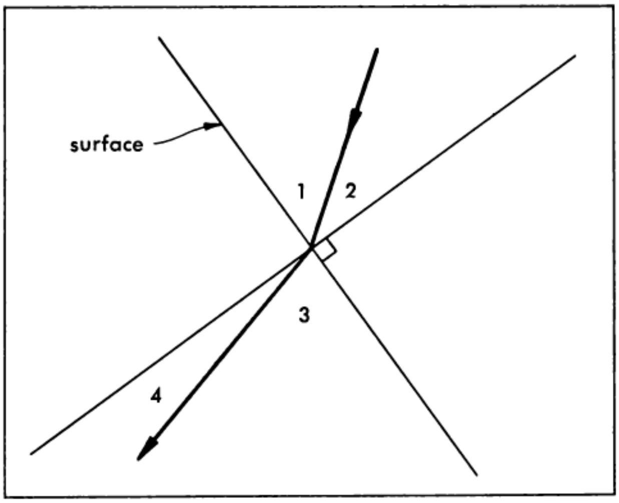
(d) If the shortest customer has eyes at a height of 5 ft. 0 in., what is the maximum allowable height of the bottom of the mirror from the floor if the customer's feet are to be visible?

7. The rearview mirror of a car is so placed that its upper and lower edges are horizontal and its center is at the same level as the center of the rear window. The driver's eye is also at this level, and the line of sight from his eye to the center of the mirror makes an angle of 30° with the line joining the centers of the mirror and the window. (See Fig. 4 – 26.) The distance from his eye to the mirror is 2.0 ft., and that from the mirror to the window is 8.0 ft. What is the least width of the mirror that is needed if the entire width (3.0 ft.) of the rear window is to be seen?



4–26 For Question 7.

8.* Which is the angle of incidence in the drawing in Fig. 4 – 27? Which is the angle of refraction?



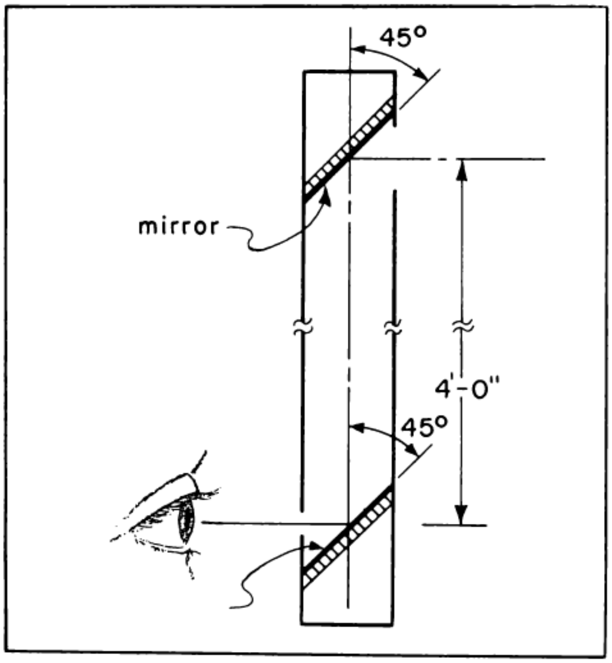
4–27 For Question 8.

9. A periscope can be made by mounting two plane mirrors at the ends of a tube. (Fig. 4 – 28.) The mirrors face each other, are parallel, and each makes an angle of 45° with the axis of the tube. An eyehole is cut in the tube opposite the center of one mirror, and a larger hole is cut in the other end of the tube, opposite the other mirror. Suppose that the distance between the two mirrors is 4 ft. 0 in., and that you are using the periscope to look over a fence at a man 6 ft. 0 in. tall who is 50 ft. away.

   (a) What is the smallest possible height of the hole in the top of the periscope?

   (b) What is the smallest possible size of the top mirror?

   Hint: Draw rays from the man to the eye, and remember the properties of similar triangles.



4–28 For Question 9.

10.* Compare Fig. 4 – 10 with the upper right-hand photo of Fig. 4 – 11. In which is the light bent more?

11. Can you give a reason why the incident ray, refracted ray, and the normal should all be in the same plane? Can you imagine a material in which this would not be true?

12.* Figure 4 – 29 shows the path of light traveling from air into glass. Is the glass on the right or the left in the drawing?

13.* In the nearly straight portion of the curve for water (Fig. 4 – 12), what is the approximate relation between *r* and *i?*

14.* In Fig. 4 – 30, what is the index of refraction for light going from air into substance *X?*

15.* Check Fig. 4 – 10. Do the incident and refracted pencils correspond to what they should be if the refractive material really is water? *Must* it be water?

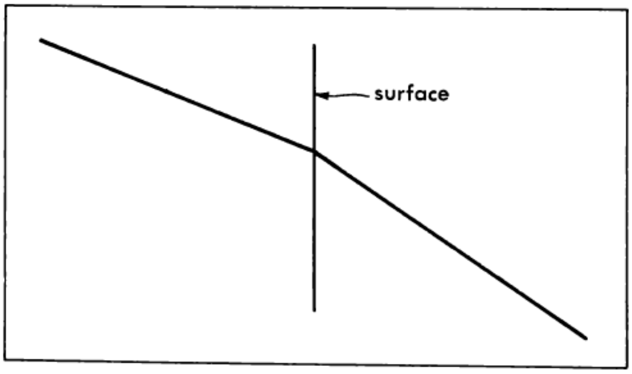
16.* For a given (nonzero) angle of incidence, which substance listed in Table 3 causes the greatest change in direction of light entering it?

17. A rectangular tank 8 cm deep is filled with water. A light ray enters the top surface of the water at a point just touching the side of the tank. After refraction it falls on a point on the bottom of the tank 3 cm from the same side of the tank.

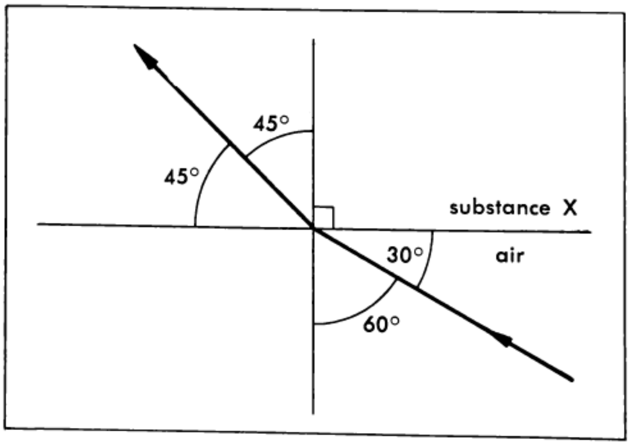
   (a) What is the sine of the angle of refraction? What is the angle of refraction?

   (b) What is the sine of the angle of incidence? What is the angle of incidence of the entering ray?

   (c) Suppose that the same tank were filled with a liquid other than water and you found that in order for the entering ray to fall on the same



4–29 For Question 12.



4–30 For Question 14.

point 3 cm from the side, its angle of incidence had to be 31°. What is the index of refraction of the liquid?

18. A narrow pencil of light enters the top surface of water in a rectangular aquarium at an angle of incidence of 40°. The refracted pencil continues to the bottom of the tank, striking a horizontally placed plane mirror which reflects it back again to the surface, and it is again refracted as it emerges into the air.

(a) What is the angle between the incident ray entering the water and the refracted ray emerging from it?

(b) If the water in the tank is 10 cm deep, what is the distance between the points on the water surface where the ray enters and where it emerges?

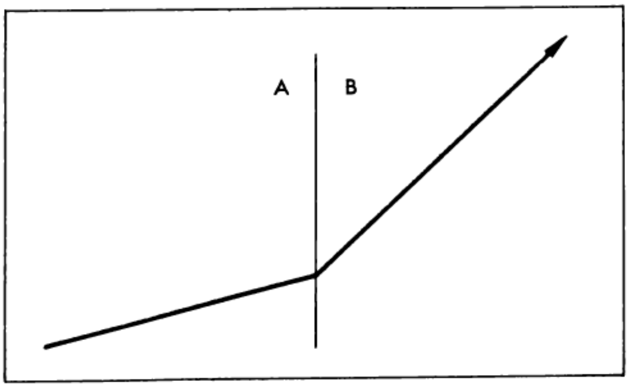
19.* Sketch one of the photographs of Fig. 4 – 11 and show by arrowed rays how the picture would look if the light were sent from below.

20. If the angle of incidence were changed in Fig. 4 – 15, would the two rays corresponding to the ones at the top of the photo still be parallel to each other?

21. Light from the setting sun comes through the earth's atmosphere along a curved path to your eye, so that the sun looks higher in the sky than it really is. How do you explain this? Illustrate your answer with a diagram.

22. As the moon moves across the sky, it passes in front of stars. Compare what is seen as the moon passes in front of a star with what you would see if the moon had an atmosphere like that of the earth.

23.* In Fig. 4 – 31, is the relative index for light going from medium *A* into medium *B* greater or less than 1?



4–31 For Question 23.

24. If the *relative* index for light going from glass into diamond is 1.61 and the *absolute* index of glass is 1.50, what is the *absolute* index of diamond?

25. If the relative index for light going from oleic acid into water is 0.91 and the index of water is 1.33, what is that of oleic acid?

26. What will you see when you look at a piece of fused quartz submerged in oleic acid? (Refer to Table 3.)

27. It is possible to place carbon disulfide, water, and kerosene in separate layers in that order, since they do not mix. A container having these three liquids in layers of equal depth has on its bottom a light source which projects a pencil of light upward through the liquids at an initial angle of 5° from the vertical.

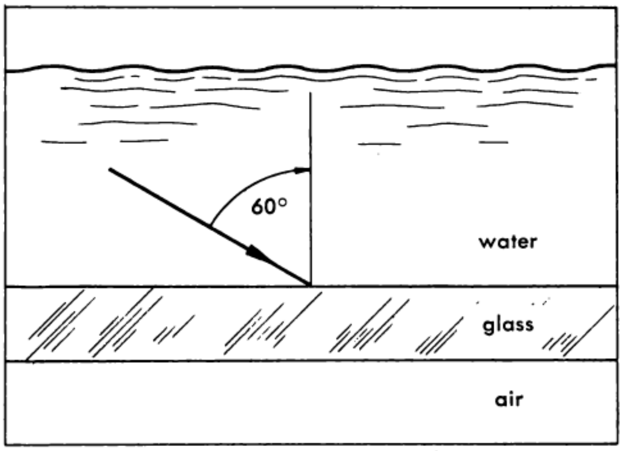
(a) Calculate the angle of refraction of the ray as it finally emerges from the kerosene.

(b) Would a pencil of light travel over the same path in the reverse direction from kerosene through water into carbon disulfide?

(c) Suppose the cylinder had been filled with water only. How would the angle of refraction of the emerging ray compare with the angle calculated above?

28.* What is the critical angle for light passing from glass to water?

29. In Fig. 4 – 32, continue the light ray until it comes out of the glass.



4–32 For Question 29.

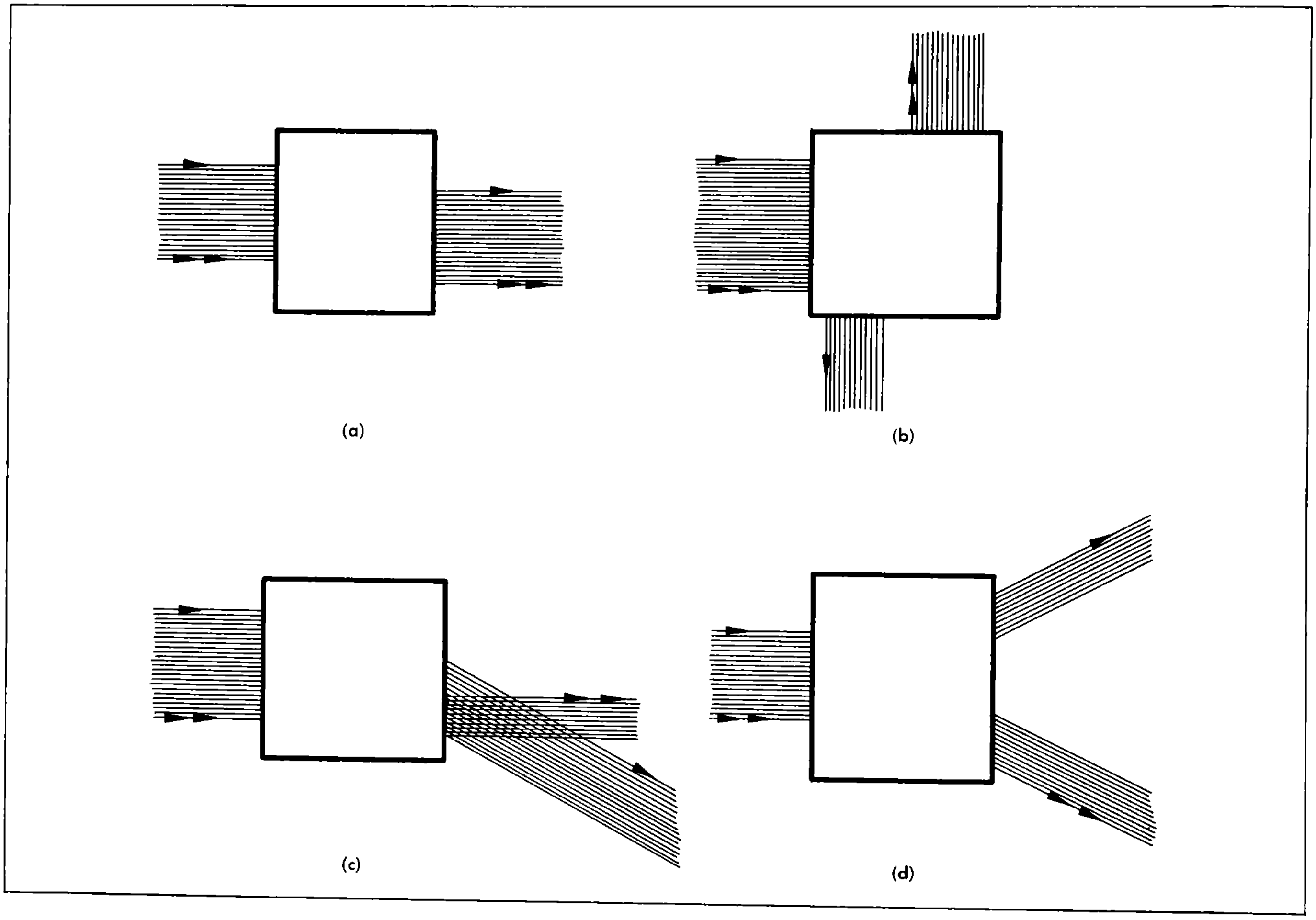
30. Carbon disulfide (refractive index 1.63) is poured into a large jar to a depth of 10.0 cm. There is a very small light source at the center of the bottom of the jar. Calculate the area of the surface of carbon disulfide through which the light passes.

31.* There are filters which pass all colors except green. There are others which pass only red. The light passing either of them appears red to the human eye. Could you distinguish between the two kinds of light by using a prism?

32. Why is it easier to observe dispersion in a prism rather than in the setup shown in Fig. 4 - 11?

33. In Fig. 4 - 33, a parallel beam of monochromatic light enters each box from the left. Draw what could be in each box to produce the effects shown. The single and double arrows on the emerging beams show the corresponding edges of the entering beam. The lines without arrows show the other edges of the emerging beam.

4-33 For Question 33.



5

The Particle Model of Light

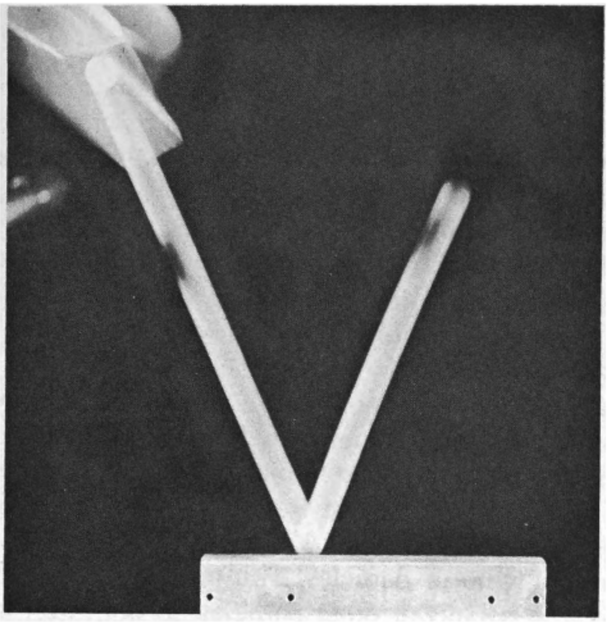
YOU PROBABLY are familiar with the atomic model, or theory, of matter. This model successfully accounts for the fundamental macroscopic properties of bulk matter. Let us now try to construct a model of light which will account for the properties of light that we have thus far studied. Like the atomic model of matter, it will allow us to draw conclusions and make predictions which can be checked by further experiments. In doing so we shall retrace ideas that were developed over several centuries. The main architect of the model we shall study in this chapter was Sir Isaac Newton.

We have seen that light always has its source in some luminous body and that it travels in essentially straight lines. Any model of light must, then, include something that starts from the luminous body and moves along a straight path. The simplest thing that we can imagine traveling in this way is a *particle,* such as a baseball. We might suppose that a luminous body gives off a stream of particles. You may object to this model on the grounds that baseballs or other particles do not travel in straight lines; they move in curved paths that bring them down to the earth. We do know, however, that the path of a baseball curves less and less as its speed is increased. We can expect that the paths of light particles which travel at a speed of 3×10^8 m/sec will hardly be curved by the earth's gravitational pull. With the particle model, therefore, we have no trouble in accounting for the propagation of light in straight lines.

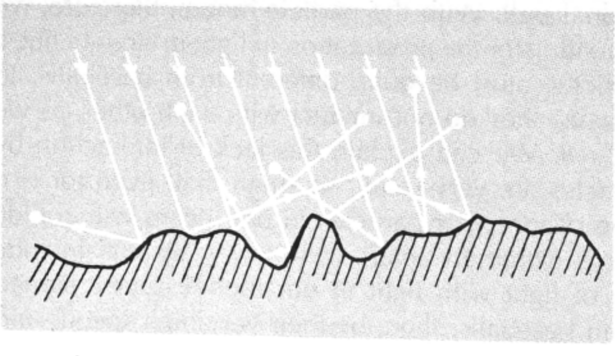
Of course, light particles must be quite different from baseballs. In particular, unlike baseballs, they do not interact with each other, as we found out in Section 3 – 8. We can explain this lack of interaction by supposing that the particles are very small, so small that, even for two intense light beams, the chance that particles in one beam will collide with those of the other is extremely small. In this way we can include the lack of interaction of light with light in our model. Our assumed light particles differ from baseballs, then, by their very high speeds and their very small size. If we keep these differences in mind, we can try to predict the properties of streams of light particles by studying the behavior of particles such as baseballs or marbles.

5–1 REFLECTION

We know that light is reflected when it strikes a surface. Do particles such as steel balls behave in the same way? To answer this question, we merely need to throw some ball bearings onto various surfaces. The balls thrown onto a clean, smooth steel plate bounce off regularly at an angle of reflection approximately the same as their angle of incidence (Fig. 5 – 1). The paths of incidence and reflection are in a plane normal to the surface; also, the speed after collision with the surface is about as great as before. If either the balls or the smooth surfaces are less resilient than steel, reflection does not take place with equal angles of incidence and reflection. Also, in such cases the speed of the balls after reflection is less than before—unlike light, which does not change its speed on reflection. To account for the specular reflection of light from smooth surfaces, therefore, we have to assume that light particles reflect like ideal ball bearings bouncing off resilient surfaces. By restricting our model to such elastic particles and using only surfaces which are elastic for the particles of light, we can explain specular reflection.



5–1 A time exposure of a ball being reflected from a steel plate. The angle of reflection equals the angle of incidence.



5–2 The diffuse reflection of particles. This occurs when a surface has irregularities larger than the size of the particles.

We have now seen that our particle model can account quite satisfactorily for the specular reflection of light. It can also account for the fact that light is sometimes reflected diffusely. Figure 5 – 2 indicates how. Individual particles bounce off with equal angles of incidence and reflection; but if the surface is uneven, the reflected particles will be traveling in a variety of directions so that the surface as a whole scatters the light diffusely. We might ask just how smooth a "smooth" surface must be to show specular reflection on a large scale. The answer is that the bumps in it must be very small compared with the size of the particles that strike it. A surface that appears quite smooth when ball bearings are bounced from it may be very rough and uneven when the small light particles are reflected. We can therefore see why metals must be highly polished before they become good mirror surfaces.

5–2 REFRACTION

Can our particle model of light account for refraction as well as for reflection? In refraction the light particles are shoved off their path in air and onto a new path as they enter the refractive material. Can we design an experiment in which moving steel balls are shoved so as to change their path in the same way?

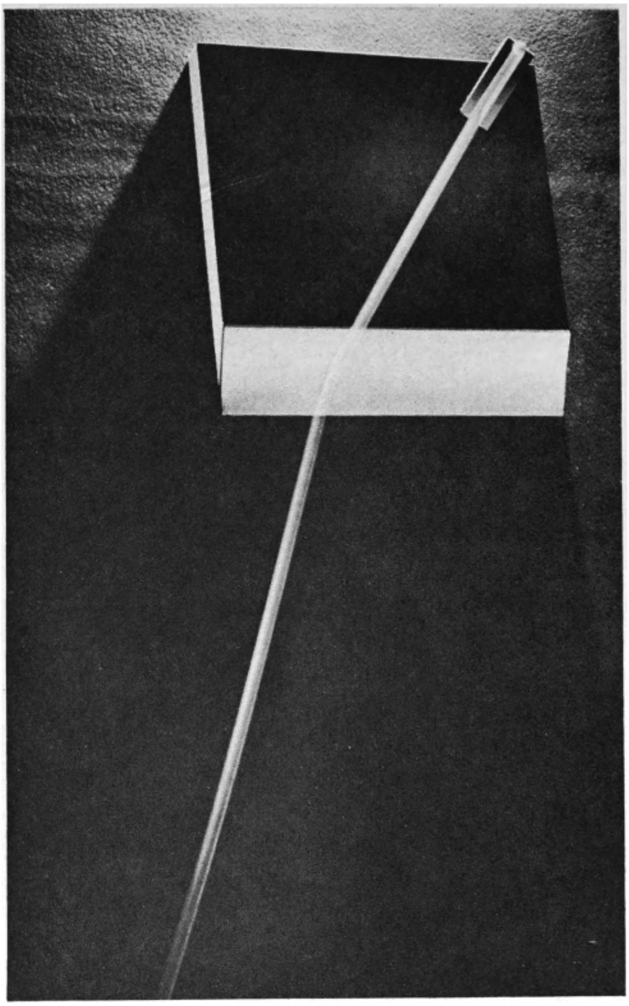
Suppose a ball rolls along a level surface and then down a steep slope onto a lower level surface. While the ball rolls on the upper level surface, it moves in a straight line at a constant speed. This upper surface, let us assume, corresponds to the region of air through which a particle of light moves in a straight line at a constant speed. When the ball comes to the steep slope, it is shoved in a direction perpendicular to the edge of the slope. The sloping region, therefore, can model the surface region of a refractive medium such as glass. In this surface region in the glass, the particles of light are supposed to be given a shove perpendicular to the boundary and toward the inside of the glass. To find out what happens to the particles of light as they pass through this region we need only look at the photograph in Fig. 5 – 3, which shows what happens to a ball as it rolls down the slope. There we see that the ball receives an impulse in the direction perpendicular to the edge, and the direction of its path is changed. If the ball rolling along the upper surface meets the slope at an angle, it rolls out from the slope along the lower surface in a different direction, closer to the normal. In the same way, light passing through a surface region in which it is shoved toward the inside of the refractive material should speed up and bend toward the normal as it enters the refractive material. The lower level surface on which the ball rolls corresponds to the inside of the refractive material. Here again the ball goes at a constant speed in a straight line like light moving inside of a piece of glass.

We can use the rolling-ball model to investigate the relation of the angles of refraction to the angles of incidence. If refraction is caused by a shove that occurs when the light passes from the outside to the inside of a refractive medium, then the relation that we find in the rolling-ball model should be the same as the relation for light. To carry out this investigation in the rolling-ball model, we always roll the balls across the upper plane at the same speed, because in a vacuum the light particles always travel at the same speed. Measurements of many pairs of angles show that the refraction of the balls agrees with Snell's law

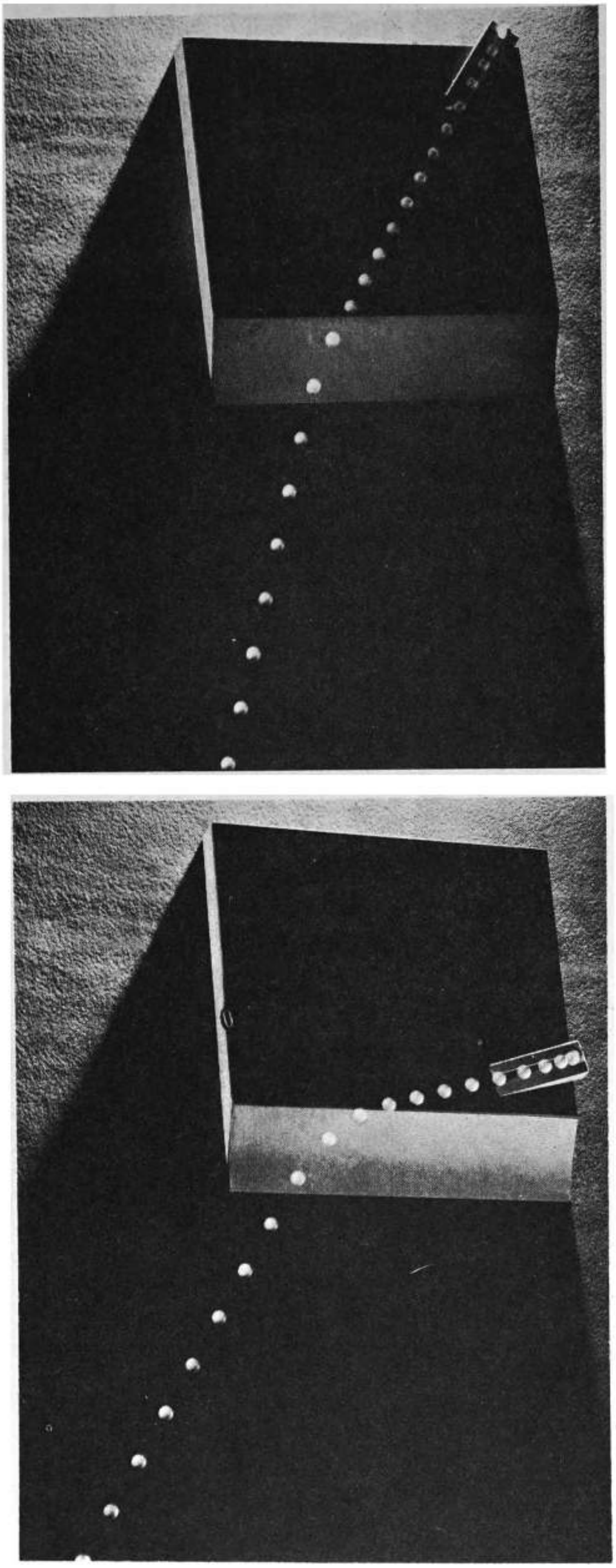
$$\frac{\sin \theta_u}{\sin \theta_l} = \text{constant}$$

where θ_u is the angle between the path of the ball on the upper plane and the normal to the boundary and θ_l is the corresponding angle on the lower plane. What is more, the constant in Snell's law depends on the difference in height between two plane regions and on the speed with which the balls move. The greater the difference in height, the greater is the index of refraction; the greater the original speed, the smaller the index. By always using the same original speed, we can model materials of different indices of refraction by planes of different heights. For example, to model the refraction of light going from air into water, the height of the sloping region must be smaller than the height needed to model the passage of light from air into glass. We find in each case that the rolling-ball model reproduces Snell's law; and by adjusting the heights between planes, we can match any index of refraction.

In discussing the rolling-ball model we do not mean that light is made of rolling balls which fall down hills. The point of this model (which was devised by Sir Isaac Newton to explain the laws of refraction) is that refraction can be explained if at the surface of every refractive material light particles are given a shove. This



5–3 A time exposure of a ball bearing rolling from a higher to a lower surface. The change in direction of the ball shows "refraction."



5-4 Two stroboscopic pictures of the ball bearing shown in Fig. 5-3. The speed on the upper level is the same in both pictures, since the ball bearing is released from the same height each time. Notice that the speeds on the lower level are also the same, although the angles of incidence are different in the two pictures.

idea is plausible. Far inside of a piece of glass or other refractive material, a light particle is surrounded by the same material on all sides. It is pushed in *no* direction; but at the surface the situation is obviously different, and a push or pull toward the inside might well occur. Then, if the push acts like that in the rolling-ball model, refraction is successfully explained.

So far this model is successful. We therefore examine it further. We can find the relation between the speeds of a ball on the upper and lower levels by taking stroboscopic pictures (Fig. 5 – 4). Analyzing many paths of many balls hitting the slope at different angles shows that, for the same v_{u} on the upper plane, the ball always attains a definite speed v_l on the lower plane. The angle of incidence on the boundary does not influence v_l. Furthermore, measurements made on many stroboscopic photographs like those in Fig. 5 – 4 show that $v_{\mathrm{u}} \sin \theta_{\mathrm{u}} = v_l \sin \theta_l$ or

$$\frac{\sin \theta_{\mathrm{u}}}{\sin \theta_l} = \frac{v_l}{v_{\mathrm{u}}}$$

Snell's law therefore holds, and the index of refraction must be $\frac{v_l}{v_{\mathrm{u}}}$.

We can also roll balls on the lower plane toward the hill starting at the speed v_l. Then we find, as we should expect, that everything is reversed. The balls arrive on the upper surface with v_{u}, and Snell's law holds in reverse just as it does for light.

We now have a particle model of refraction, and — as usually happens when we have a model — it suggests many experiments. For example, our model says that the speed of light in a refractive material is independent of the angle of incidence and greater than the speed of light in vacuum. We should make measurements of the speeds of light in refractive materials and in particular find out if they agree with

$$\frac{v_{\mathrm{m}}}{c} = n_{\mathrm{m}}$$

where v_{m} is the speed in the refractive material, c the speed of light in vacuum, and n_{m} the index of refraction. We shall discuss such measurements in Section 5 – 7. Also we should check to see that the speed of light in vacuum does not depend on the prior history of the light. According to our model, a light particle slows down in going from a refractive material into vacuum by just as much as it speeds up in going from vacuum into the material. The constancy of the speed of light in vacuum—regardless of where the light comes from—is a check on this particle model of refraction.

5–3 SOURCE STRENGTH AND INTENSITY OF ILLUMINATION

In order to find the ratio of the distances to two similar stars in Chapter 2, we used the inverse-square law relating the intensity of illumination to the distance by $I = \frac{k}{r^2}$, where r is the distance of the source of light and k must have something to do with the strength of the source. Let us now investigate the intensity of illumination and the strength of sources a bit more closely.

In order to measure the intensity of illumination, we can use a photoelectric cell or other light meter, such as those used by photographers. These light meters have a pointer which moves across a scale. Higher scale readings on the meter correspond to greater illumination as we estimate it by eye. Our first job is to discover a way to calibrate the meter by a method of our own that we understand.

We start by finding a set of equal sources of light. By "equal" we mean that the light meter registers the same reading for each source when placed by itself at a set distance r_0 in front of the light meter in an otherwise dark room (with dark walls to eliminate reflections). We conclude that each light source is causing the same number of light particles to fall on the sensitive element of the light meter in one unit of time, say a second.

The particle theory of light now suggests that if we place two of these equal sources at the same distance r_0, twice as many particles of light will illuminate the light meter per second. Indeed, using any two of our standard light sources, each a distance r_0 away from the meter, we always get the same pointer reading on the meter; this reading indicates greater illumination than we get with only one standard source at the standard position.

Let us make our calibration of the scale for the light meter by marking the position of the pointer when one standard source is at the standard distance r_0. We put the number 1 right on the scale at that point. We put the number 2 on our scale at the position of the pointer when two standard sources are at the standard distance. We can do the same thing with three standard sources at the standard distance in front of the meter, and so on. We have now calibrated our light meter to read illumination in multiples of the illumination given by one standard source placed at the distance r_0 in front of the meter. When we use the meter to measure the illumination of a standard source at a distance different from r_0, our scale tells us how great that illumination is in terms of the number of our standard sources that would have to be placed the distance r_0 in front of the meter to give the identical effect.

Suppose that we originally chose the distance r_0 in front of the meter to be very long compared to the size of our light source and also long compared to the dimensions of the sensitive element of our light meter. We can now take one of the standard sources and move it to distances in front of the light meter which are smaller than r_0. Thus, we can find out how the intensity of illumination at the sensitive element of the light meter changes as we change r, the distance from the source to the meter. The results of such an experiment are shown in Table 1.

The third column shows the product of the intensity and the square of the distance. This

TABLE 1

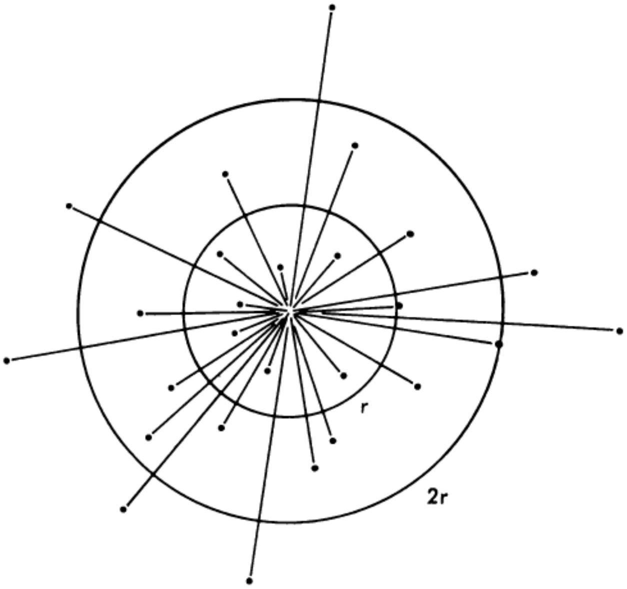
r (M)	I	Ir^2
1.20	1.0	1.4
1.00	1.5	1.5
.80	2.3	1.5
.70	3.2	1.6
.60	4.3	1.5
.50	5.4	1.4
.40	8.5	1.4
.30	17.7	1.6
.25	23.5	1.5
.20	33.3	1.3

The intensity of illumination at different distances from a source.

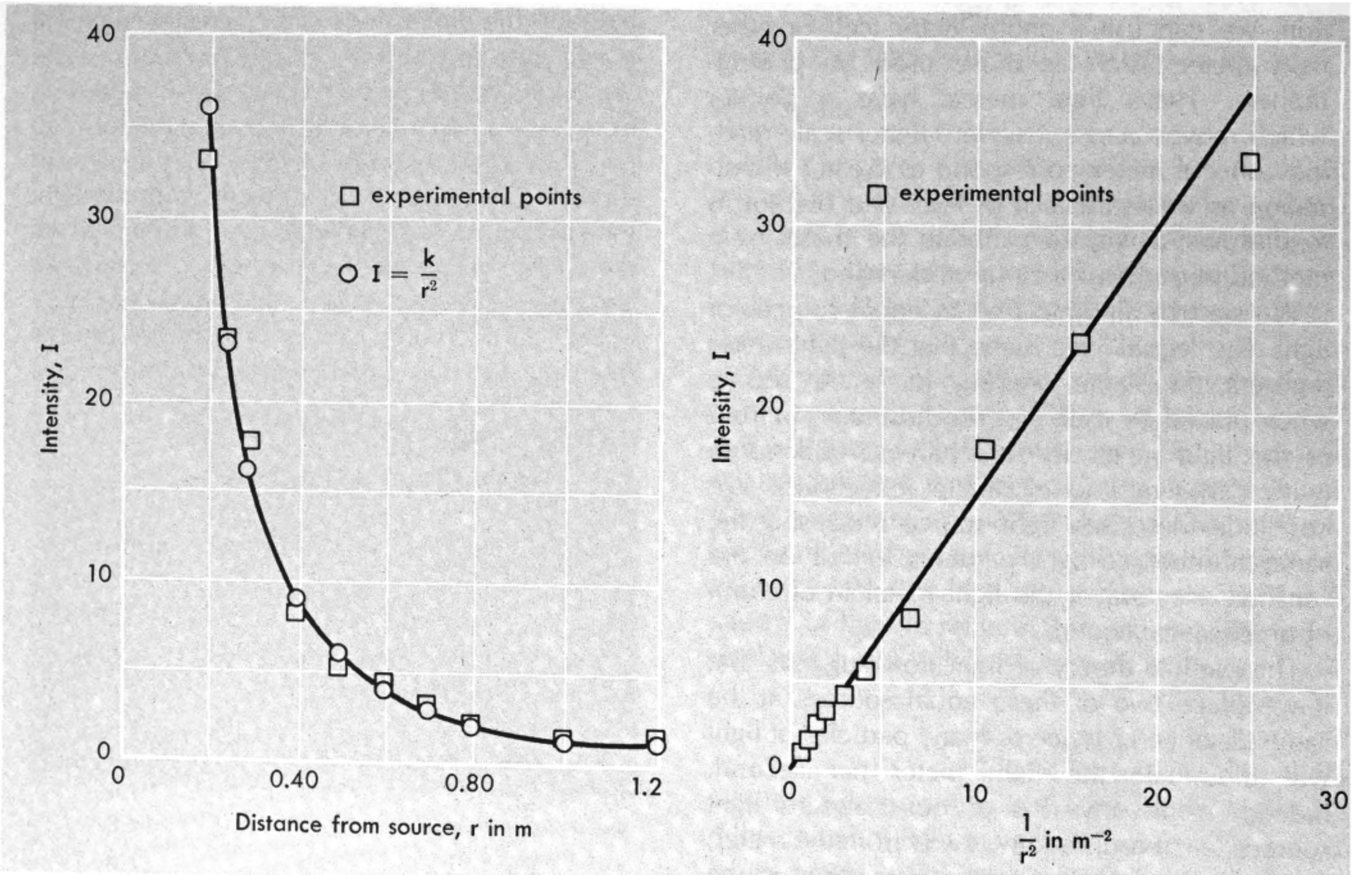
product is almost a constant, as we should expect if the inverse-square law is true.

The data from Table 1 are plotted in Fig. 5 – 5.

The result that the illumination varies inversely with the square of the distance from the point source is just what we would expect according to the particle theory of light. We suppose that illumination is directly proportional to the number of particles hitting a unit area in one second. The particles move outward in all directions from the source, traveling in straight lines, and we assume that none of them are lost as they travel. At a distance *r* from the point source, the particles will be spread over a sphere of radius *r* and of area $4\pi r^2$. (See Fig. 5 – 6.) At a distance $2r$ from the source, the same total number of particles must be spread over an area $4\pi(2r)^2$ or $16\pi r^2$. Thus the number of particles per unit area at $2r$, and hence the illumination at $2r$, will be only one-fourth as great as the number of particles per unit area, or illumination, at *r*.



5–6 The particles which are spread over the area $4\pi r^2$ at radius r will be spread over the area $16\pi r^2$ at the radius 2r. Thus the intensity will be only one fourth as great at distance 2r as it was at distance r from the point source.



5–5 The graph on the left shows the intensity of illumination at different distances from a source. The experimental values agree closely with points computed, assuming an inverse-square relation. The graph on the right, where intensity is plotted against $1/r^2$, shows close agreement with a straight line.

This theoretical derivation of the inverse-square law is the same as that given in Chapter 2, but in place of a vague amount of light passing through greater and greater areas, we now have a certain number of individual light particles.

When the source and the receiver of light particles are close together compared to their sizes, the light particles can pass from one to the other at odd angles to the receiving surface instead of along paths which are nearly parallel to each other and perpendicular to the sensitive surface of the light meter. We are, therefore, not surprised if we find greater deviations from the inverse-square law when the source and light meter are close together.

Like the standard units of distance or time, our standard light source is an arbitrary unit of source strength. Source strengths were systematically measured long ago, so long ago that the usual unit is the candle. You can imagine the prescription for making a standard candle. This candle is to be a $\frac{7}{8}$ in., No. 6 sperm candle burning at the rate of 120 grains per hour. When it is designed to burn just this way, it should then be just as bright in France as it would be in Alaska.

Of course, as has happened to other standards, the prescription has been replaced by a new one, but the basic ideas are the same. The standard candle now is defined in terms of the light that comes through a hole 1/60 cm^2 in area, behind which is a platinum body heated to its melting temperature. The new standard candle is much more accurately reproducible than the old one, but there may be some question as to whether or not it is easier to set up your own standard source.

The standard unit of intensity of illumination is based in turn on the standard candle. It is the intensity given by one candle one foot away. This unit of intensity of illumination is known as the foot-candle.*

5–4 LIGHT PRESSURE

Whenever a baseball or a stone hits us we feel a push. The pushes exerted by the tiny molecules of a gas bumping into the walls of a container or pounding on our skin give rise to gas pressure. Do the particles of light bumping into a mirror similarly give rise to light pressure? The particle theory of light suggests that they do. Unfortunately, the particle theory we have developed so far is not specific enough to tell us how much light pressure to expect from a given stream of light particles. For example, we do not know what intensity is given by a mole of light particles passing each second through a square meter.

But even without this detailed knowledge we do know that if light exerts a pressure at all, it must be very small for an ordinary beam of light. We do not knock over a feather by switching on a bright light. Even experiments with delicate apparatus failed to reveal the existence of light pressure until the beginning of our own century, when Peter Lebedev in Russia and Nichols and Hull in the United States succeeded in finding and measuring this pressure. From their experimental results we know that the pressure of bright noon sunlight is just about the same as the pressure from the weight of a monolayer of oleic acid lying on a horizontal surface. This is a very small pressure, but not all light pressure is so small. Experiments verify the prediction of the particle theory that the pressure of light increases with the intensity of illumination. At the surface of our sun, for example, light pressure is several orders of magnitude larger than the pressure of sunlight here, and at the surfaces of some other stars the light pressure is enormous.

5–5 ABSORPTION AND HEATING

A good reflector of light does not warm up appreciably when placed in sunlight, whereas a dark, absorbing surface can become quite warm when exposed to the sun. Dirt, shoveled up from a sidewalk along with snow after a snowstorm, melts its way down through the snow when the sun comes out. The black dirt, a good absorber, gets warm and melts the snow beneath it faster than the white reflecting snow melts by itself.

Can the particle model account for this behavior? When particles are neither reflected from a material nor transmitted by it, they must come to rest. When a piece of matter is hit by

---

*You must read this unit as a single word; it does *not* mean feet × candles. For example, at 1 ft. 3 standard candles give 3 foot-candles illumination; at 2 ft. they give (3 foot-candles) $(1\text{ ft}/2\text{ ft})^2 = \frac{3}{4}$ foot-candle.

particles that neither pass through it nor bounce off it, does the material heat up? We can easily show that heating really occurs with particles by substituting a hammer for the particles and using a block of lead for the absorbing material. If we strike the block a series of rapid blows (like a series of particles striking the lead), we find that the hammer does not bounce back (the particles do not reflect) and the lead gets warm. As a check to see that this heating is not an effect we get with every particle bombardment, let us repeat the experiment using a heavy steel block in place of lead. The hammer now bounces back with every blow (it reflects), and the steel does not noticeably warm up. If light particles act like particles of ordinary matter, we would expect them to heat any material in which they are stopped. The particle model therefore agrees with the observed heating of materials which absorb light and with the fact that materials which reflect or transmit all the light do not heat up.

Absorbing materials are not usually black. They reflect some of the light which strikes them, and they may transmit some of it. In order to make a model of a substance which reflects some particles specularly while it absorbs other particles from the same light beam, we may imagine that the material contains small regions that stop light particles, producing heat, intermingled with regions that reflect light particles and produce no heat. Even the best reflectors, such as mirrors, warm up a little when placed in the sun, and we conclude that a mirror surface must contain a few scattered absorbing regions. A transparent substance such as a pane of glass must also contain some absorbing regions since we observe a slight decrease in the light intensity of light emerging from glass inserted in a light beam. Furthermore, as expected, the glass does become slightly warm.

5–6 SOME DIFFICULTIES WITH THE PARTICLE THEORY

We have found that we can make a particle model which accounts for specular and diffuse reflection, which agrees with Snell's law of refraction, and which suggests the inverse-square law for the diminution of the intensity of illumination with distance. The model also suggests that light should exert a pressure and that heating should be associated with light absorption. These last two effects, which we had not previously discussed, are found experimentally; and they may serve as examples of the value of a model in leading us to new investigations. The model is successful in describing coherently much that we already know by giving us a picture into which our knowledge fits. It also suggests that other things should take place about which we may have had no idea. It gives us the clue to what may otherwise be unexpected, because, like the pressure of light, the unexpected may exist in the model. We can then check the model by comparing nature with what the model predicts.

So far with our model of light we have been very successful. Are there any respects in which the particle model is less successful? Are there aspects of the behavior of light which are poorly described by the model, or even in conflict with it?

Let us take a closer look at what happens when light is refracted. Although we know how to account for refraction and reflection separately on the particle model, can we account for the fact that both take place when a light beam reaches the surface of a refractive material? To produce specular reflection the surface shoves the particles of light in one direction; to produce refraction on our model it shoves them the opposite way. How does it decide which way to shove? This question bothered Newton, and he attempted to explain the partial reflection and partial refraction of light by supposing that the particles of light had properties which changed periodically with time. What happens to the particle then depends upon the time at which it hits the surface.

Another attempt to explain the splitting up of light beams might be made by imagining that the surface is divided into little reflecting regions and little regions that refract. Either of these extensions of our model might account for a definite ratio of reflection and refraction. But they are in trouble because the ratio of refracted to reflected light decreases as the angle of incidence increases. Perhaps the difficulty can be overcome, but the model is getting complicated. When we have to fix it up too much, we look on it with suspicion.

5–7 THE SPEED OF LIGHT AND THE THEORY OF REFRACTION

According to our particle theory, the speed v_m of light in a refractive material must be greater than the speed c in vacuum. In Section 5 - 2, we found $v_m = cn_m$, where n_m is the index of refraction. This is a quantitative result of the model which we can test if we can measure the speed of light over a short distance—in a tank of water, for example.

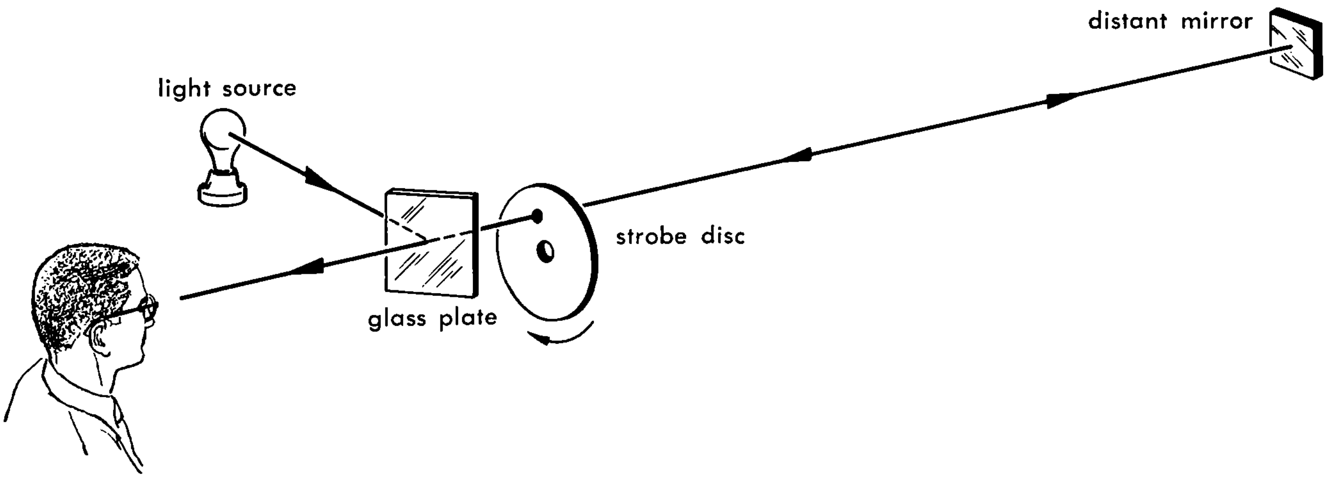
In Newton's time such a measurement was not possible. The first successful measurements of the speed of light over a measured path on earth were made by Fizeau (1849). His method of measurement is based on the same idea as that of the stroboscope. Think of directing a beam of light at the hole in a stroboscope. Little pulses of light are let through each time the opening in the whirling disc passes the beam. Now let these pulses strike a mirror so that they are reflected back to the stroboscope again (Fig. 5 - 7). On the return trip a pulse can pass through the stroboscope only if the opening in the disc is again in the right position.

For a stroboscope with a single opening in the disc, the disc must rotate at least one full turn in the same time that the light goes to the mirror and back. By measuring the speed of rotation when light can pass both ways, we can time the light as it goes the known distance from the stroboscope to the mirror and back again. Thus we determine the speed of light.

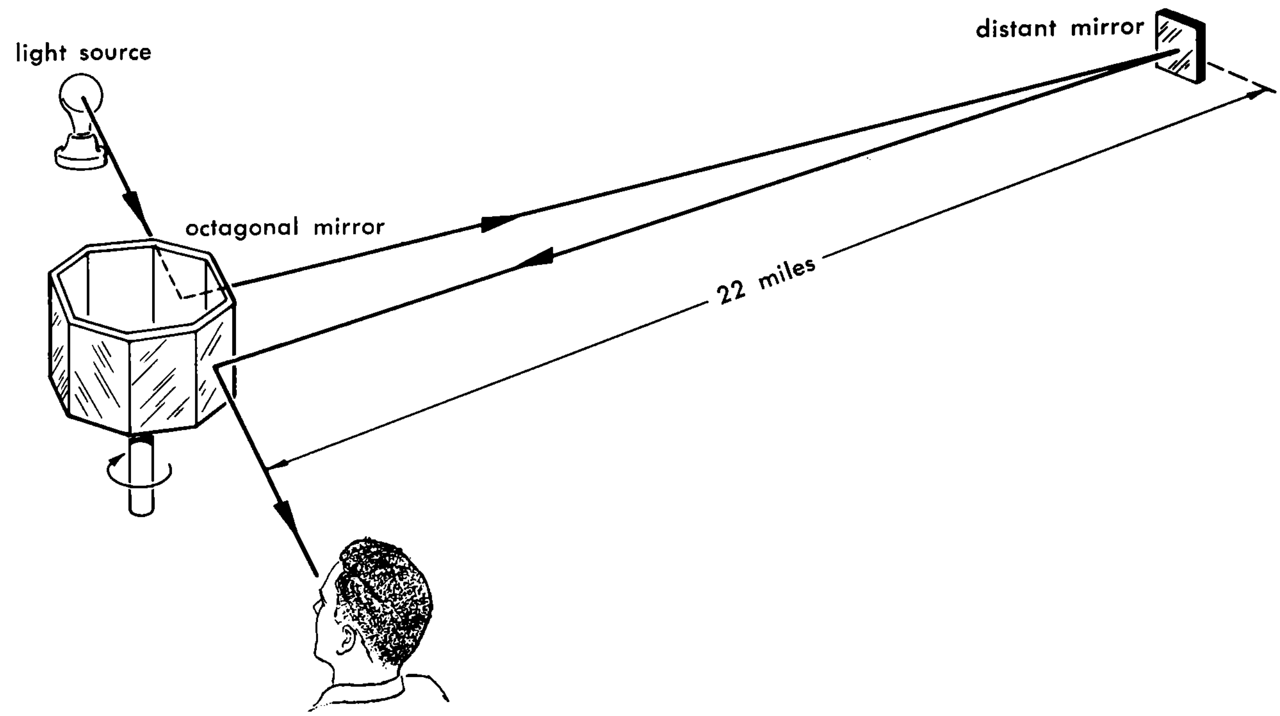
Of course, no ordinary stroboscope will turn fast enough. But, by designing a special high-speed stroboscope with many holes, Fizeau succeeded in measuring the speed of light in air. Other physicists improved Fizeau's method, among them the French physicist Cornu. His stroboscope had 200 openings and could turn at 54,000 revolutions per minute. As he increased its speed from zero to the maximum he measured twenty-eight successive eclipses and brightenings of the light coming back through the disc from a mirror 23 km away. By using a rotating mirror to replace the rotating disc of the stroboscope, Foucault in France and later Michelson in the United States increased the accuracy of this method still further (Fig. 5 - 8). Michelson was able to determine the speed of light to an accuracy of about 3 km/sec (1 part in 10^5).

For this accurate measurement Michelson used a long path. But the same method can be used over short paths to find the speed of light in liquids, such as water, though with less accuracy. In 1862, Foucault measured the speed of light in water and found it to be close to 2.23×10^8 m/sec. This is almost exactly three-fourths of its speed in air. Some years later, Michelson determined the speed of light in carbon disulfide, a liquid with an index of refraction of 1.63, and found it to be 1.71×10^8 m/sec.

When it encounters these results our particle theory of light really strikes a snag. In order to account for Snell's law, we were forced to assume that the speed of light is higher in refrac-



5–7 Timing a light pulse with a stroboscope. Light from the source reflects from the glass plate and goes through the hole in the strobe disc to a distant mirror. The returning pulse passes through the strobe disc and through the glass plate to the eye of an observer.



5-8 Michelson reflected light from an eight-sided rotating mirror. With this equipment he made precise measurements of the speed of light.

tive materials than in vacuum. Using Newton's particle model, we obtained $v_{\text{water}} = n_{\text{water}}c$. Consequently, on the model the speed in water should be about $\frac{4}{3}$ of that in air, and in carbon disulfide the speed should be about $\frac{5}{3}$ of that in air. The observed values are almost exactly $\frac{3}{4}$ and $\frac{3}{5}$ of the speed in air. Not only have the experiments failed to show the ratios that we expected—they indicate that the ratios are inverted. Our particle model of light appears to fail.

5-8 THE STATUS OF THE PARTICLE MODEL

In this chapter we tried to develop a model of light based on the behavior of particles. This attempt was successful as long as we chose only *some* of the optical phenomena we know about, and as long as we did not always demand precise and quantitative agreement between the predictions of the theory and the results of experiments. When we included all of the phenomena and asked for quantitative checks, however, the model failed in several important respects. Since our particle theory of light failed, it must be either abandoned or modified.

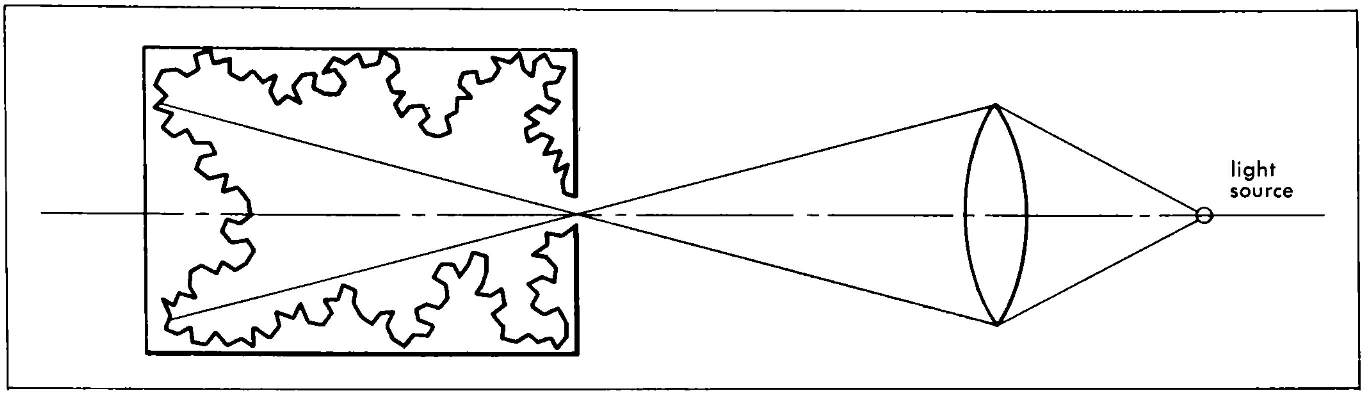
We are now in a situation in which physicists often find themselves. From time to time, a theory which successfully connects a whole group of experimental results and observations fails to account for a new observation. Then we must modify the theory, or start over again and try to invent a new one. It is almost always true, however, that the efforts spent on the old model have not been wasted. The theory has served two purposes: it has shown that some of the things that were observed can be related to each other; and it has suggested new experiments. The particle picture of light, which here appears unsatisfactory, has played a very important part in the history of physics, and it still contributes to our understanding of light. We shall have more to say about it later. For the moment we put it aside and try a different model.

QUESTIONS

1.* Why should a Ping-Pong table have a more even surface than the floor of a basketball court?

2. Consider the box shown in Fig. 5 - 9, whose inside walls are perfectly reflecting. What do you predict will happen when you shine a beam of light into the box? (Use the particle model of light to arrive at your answer.)

3.* In Fig. 5 - 4, if you keep the speed of the ball



5-9 **For Question 2.**

bearing on the upper level constant by launching the ball always in the same manner, at what angle of incidence to the hill will you get the highest speed on the lower level?

4.* What does the experiment of particle model refraction suggest about the speed of light in a material, that was not apparent in our refraction experiments with light?

5. The index of refraction of carbon disulfide is 1.63. What should the speed of light be in this liquid, according to the particle model of refraction given in Section 5 - 2?

6. The speed of light measured in a certain medium is 2×10^8 m/sec.

   (a) According to the particle model of light, what should be the index of refraction of this medium?

   (b) Using the index of refraction from part (a), what would be the path of a light ray which passes from air into the medium at an angle of 30° from the normal to the surface of the medium?

7. Occasionally there occurs in the heavens an explosion of a star, producing what is known as a super nova. The star suddenly becomes many times brighter than before. As you know, the stars are so far away that the light from them takes many years to reach us. An explosion that we observe must have occurred a long time ago, and the light has been traveling toward us ever since. We see the explosion as a bright white light, not as a series of different colors arriving at different times.

   (a) What does this show about the speed of light of different colors in vacuum?

   (b) Try to suggest a particle model for dispersion in prisms consistent with the single speed of light of all colors in vacuum.

8.* How can you check several light sources to see if they are equal sources of light?

9.* Suppose a light meter gives a reading of 1 unit when it is a distance *r* from a light bulb. At what distance from the same source will the meter give a reading of 2 units?

10. Can you explain the different strengths of light sources in terms of particles of different size? How about the decrease in intensity caused by inserting a partially absorbing sheet of matter in a light beam? Can different intensities at varying distance from a source be accounted for in this way?

11. What assumptions do we make in using the inverse-square law to determine star distances?

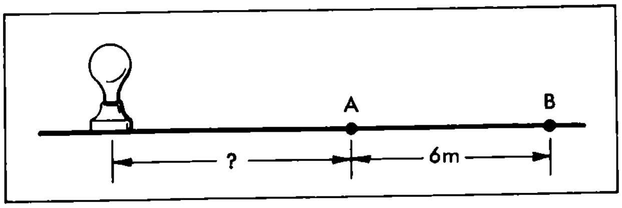
12. What does the particle theory predict about the intensity produced by an *extremely* weak light source—so weak that only a few particles are emitted per second? How might you test this prediction?

13. It is found that a 40-candle source placed 3.0 feet from a light meter gives the same scale reading as an unknown source 1.2 feet from the meter.

    (a) What is the strength of the unknown source?

    (b) What is the light intensity in foot-candles read on the meter?

14.* In Fig. 5 - 10, how far to the left of A is the light source, if it gives an intensity of 8 foot-candles at A and $\frac{1}{2}$ foot-candle at B?



5-10 **For Question 14.**

15. A light meter calibrated in foot candles is 2.0 feet from a mirror. A very small 25-candle-power source is placed midway between them. What is the reading of the meter if the mirror is
   (a) a plane mirror?
   (b) a concave mirror with a diameter of 6 inches and a focal length of 1 foot?

16. In the use of radar, a beam of radiation is sent out from a source. Some of this radiation falls on a distant object, such as an airplane, is reflected by the object, and is detected when it returns to its starting point. The radiation used behaves like light. Assuming that the source of the radiation is equivalent to a point source of light and that the object reflecting the radiation is a diffuse reflector, can you convince yourself that the intensity of the returning radiation varies inversely with the fourth power of the distance from the source to the reflecting object?

17. Upon which object would light exert the greater pressure, a clear piece of glass or a mirror?

18. A sensitive thermometer placed in the different parts of the spectrum from a prism will show a rise in temperature. This shows that all colors of light produce heat when absorbed. But the thermometer also shows a rise in temperature when its bulb is in either of the two dark regions beyond the two ends of the spectrum. How can the particle theory account for this?

19. A thermometer is placed in the beam of light coming from a lamp. Its reading increases until it becomes steady at 26° C. When a piece of ordinary window glass is placed between the lamp and the thermometer, the reading drops to 23° C.
   (a) What can you conclude about the nature of the light coming from the lamp?
   (b) If the window glass is replaced by Corex glass (see Section 3 – 5), would you expect the thermometer reading to be above or below 23° C? Above or below 26° C?
   (c) Can you be equally sure of both of your answers?

20. A calibrated light meter shows that one glass plate transmits 80 percent of the light from a fixed source.
   (a) What fraction of the light will be transmitted by two, three, etc., up to ten such plates?
   (b) Draw a graph of the fraction of light transmitted as a function of the number of plates.
   (c) Plot the logarithm of the fraction of light transmitted as a function of the number of plates.
   (d) How many plates are needed to absorb 90 percent of the light?
   (e) Be prepared to discuss the effect of reflection on the accuracy of your results.

21.* Why is it possible to see light through the hole in the strobe disc in Fig. 5 – 7 when the disc is rotating very slowly?

22. Using the data given in Section 5 – 7 about Cornu's application of the Fizeau method, compute the shortest distance from the rotating disc to the mirror that will permit the returning beam to pass through the opening immediately following the one through which it started.

23. In the Foucault-Michelson modification of the Fizeau method, rotating mirrors are used in place of the rotating disc (Fig. 5 – 8). Assuming that mirrors are rotating 500 times per second and that the distance from them to the fixed mirror is 10.0 km, show that a mirror face rotates through an angle of 12° before it reflects the beam a second time and that the angle between this reflected beam and the original direction of the light is therefore 24°.

24. It might be imagined that on each reflection the speed of light decreases slightly. Assuming that you can measure the speed of light (see Section 5 – 7), describe an experiment designed to test the assertion that reflection does not change the speed.

25. Can you modify your proposed experiment (Problem 24) to test the assertion that the speed of light in vacuum is independent of previous refractions?

26. Suppose that someone were to propose to you that sound consists of small, rapidly moving particles that are emitted by a source and that affect your ear when they strike it. What evidence could you use to support or to refute it?

27. We have said that light particles must be very small. Suggest an experiment to show that they must also have an extremely small mass.

28. Is it possible that the slowing down of light as it goes from a vacuum into a transparent material is some kind of friction effect?

29.* Why does the measurement of the speed of light in water cast doubt on the validity of the particle model as developed in this chapter?

6 Introduction to Waves

6–1 A WAVE: SOMETHING ELSE THAT TRAVELS

In the last chapter we considered at some length a particle model of light, in which we supposed that light consisted of a stream of particles, or corpuscles. We found that this model fails to provide completely satisfactory explanations for some of the behavior of light that we observed. We therefore find ourselves faced with a choice: we can try to construct a better particle model that will succeed where the earlier one failed,, or we can look for a new model based on a completely different concept. Let us try the second approach.

The most basic thing to be accounted for in *any* model of light is the fact that light travels through space. In looking for a new theory, we first ask whether there is anything except a particle (or stream of particles) that can move from one point to another. The answer is "yes." Consider, for example, what happens when we drop a pebble into a quiet pond. A circular pattern spreads out from the point of impact. Such a disturbance is called a *wave,* and if you watch the water closely enough, as such a wave moves across the surface, you will find that although the water may be churned and jostled locally, it does not move forward with the wave. This is quite clear if you watch a bit of wood or a small patch of oil that may be floating on the pond. The wood or oil moves up and down as the wave passes; it does not travel along with the wave. In other words, a wave can travel for long distances, but once the disturbance has passed, every drop of water is left where it was before.

If we look around us, we can find all sorts of examples of waves. For instance, we notice an American flag as it ripples in the breeze at the top of a flagpole. The ripples, or waves, travel out along the cloth. Individual spots on the cloth of the flag, however, hold their positions as the waves pass by. The fourth white star in the bottom line on the field of blue always remains the fourth star in the bottom line and its distances from the four edges of the flag remain unchanged. Just as the water does not travel with the water waves, so the cloth of the flag remains in place when the waves have passed through it.

Some waves are periodic or nearly so; the motion of the material repeats itself over and over. Not all waves, however, have this property.

For example, when you slam the door of a room, the air in the doorway is suddenly compressed, and this single short compression passes as a disturbance across the room, where it gives a sudden push to a curtain hanging over the window. Such a wave of short duration is called a *pulse*.

Here is another example of a wave pulse. We place half a dozen pocket-billiard balls (plastic croquet balls will work, too) in a straight line so that each ball is touching the next one. We then roll another ball so that it strikes one end of the row head-on. The ball at the other end of the row moves away at a speed apparently equal to that of the ball we rolled in. A pulse like that on the water surface or on the flag has traveled through the row of balls from one end to the other. Each ball has been disturbed; this disturbance has passed along the entire line of balls; but no ball has moved from one end of the line to the other.

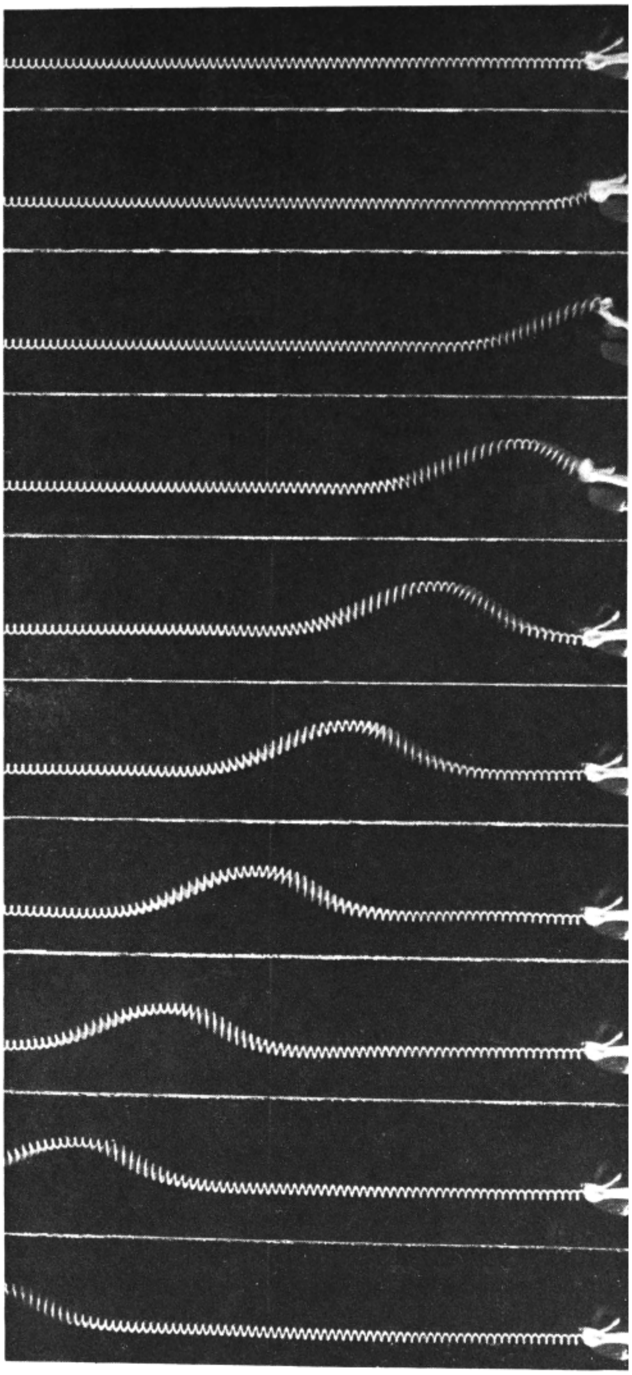
You see another kind of wave—let's call it a "wave of starting"—if you watch a line of cars that has been stopped while waiting for a traffic signal to change from red to green. Shortly after the signal changes to green, the driver of the first car starts his car forward. A moment later, the driver of the second car, seeing that the car ahead of him is in motion, starts his car forward. Still later, the driver of the third car gets his car under way, and so on down the line away from the traffic signal. You can see a "pulse" move down the line of cars. It is interesting that this "starting pulse" travels in one direction while the cars travel in the other direction.

Just how fast the starting pulse travels back depends on how fast the various drivers react and how their cars respond. If we were able to handpick a group of drivers with identical reaction times, and provide them with cars that accelerated in exactly the same way, the starting pulse would travel backward at a uniform speed.

What is alike in all of these examples? In each case the disturbance travels through some medium—through the water, the cloth of a flag, the billiard balls, or the line of cars; but the medium does not go along with the disturbance. Disturbances which travel through media are what we mean by waves. We can now answer the question we asked at the beginning of this section: is there anything except a particle that can move from one point to another? A wave, a thing which is not itself a particle of matter, can move from one place to another.

6–2 WAVES ON COIL SPRINGS

Do waves really behave like light? To find out, we must know more about them. When we know how they act, we can compare their behavior



6–1 The generation and motion of a pulse along a spring shown by a series of pictures taken with a movie camera.

with what we know about light and with other things that we can find out about it. The variety of examples we have mentioned also suggests that waves are worth studying for their own sake.

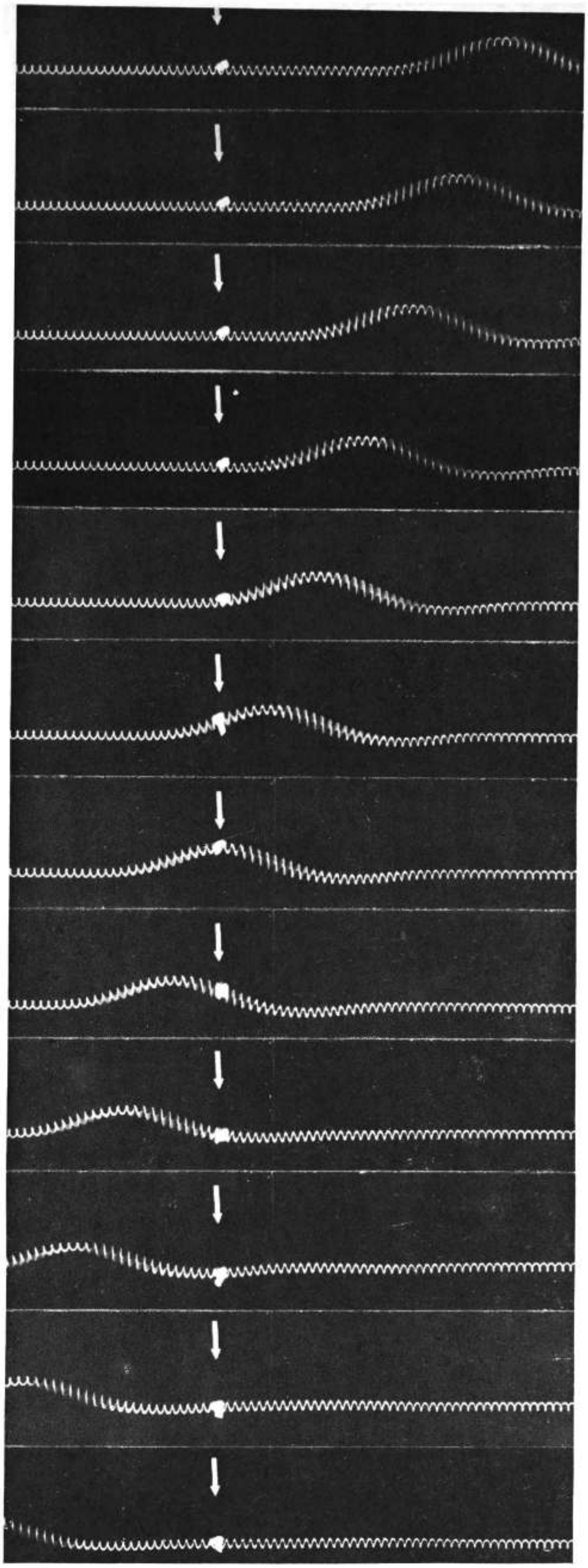
It is convenient to start our study of waves with a coil spring.* Figure 6 – 1 shows pictures of a pulse traveling along such a spring. These pictures were taken by a movie camera at intervals of $1/24$ of a second.

We see that the shape of the pulse does not change as it moves along. Except for the fact that the pulse moves, its picture at one moment is just like a later picture. Also we see that the pulse moves the same distance in each interval between pictures—it moves along the spring at constant speed.

The spring as a whole is not permanently changed by the passage of the pulse. But what happens to each small piece of spring as the pulse goes by? To help us fix our attention on one piece, we can mark the spot by tying on a bit of white string or ribbon as shown in Fig. 6 – 2. If we then shake the spring to start a pulse moving along it, we can see how the marked spot is displaced. We find that it moves at right angles to the spring as the pulse passes it.

Other pieces of the spring, as well as the marked spot, also move. We can see which pieces are moving and which way they go if we look at two pictures, one of which is taken shortly after the other. Here we shall use two successive pictures taken from Fig. 6 – 2. We have printed these two pictures together in Fig. 6 – 3 so that we see the pulse in two successive positions just as we would see it in a rapid double exposure. Below the photo in Fig. 6 – 3 we have traced the pulse in its earlier position, and the gray line shows the later position. As the arrows show, while the pulse moved from right to left, each piece of the coil in the right-hand half of the pulse moved down and each piece of coil in the left-hand half moved up.

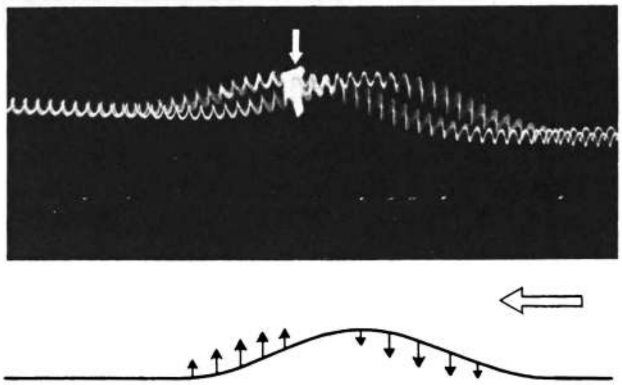
If the pulse were moving from left to right, just the reverse would be true, as we show in



6–2 **The motion of a pulse from right to left along a spring with a ribbon around one point. The ribbon moves up and down as the pulse goes by, but does not move in the direction of motion of the pulse.**

---

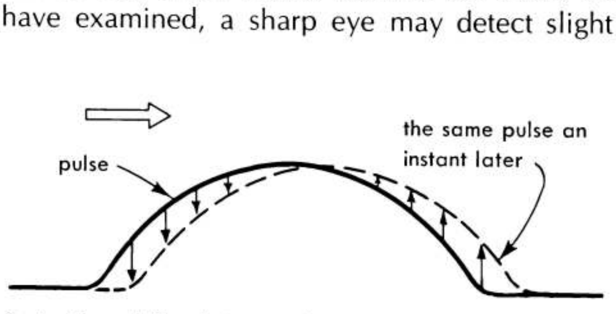
*If you find it hard to get a coil spring, a flexible clothesline or a rubber tube will also do pretty well. Tie one end to a doorknob and shake the other. If the clothesline or tube is sufficiently heavy, you will get good pulses that travel slowly enough for easy observation.



6–3 The relation between the motion of a pulse traveling from right to left and the motion of the coil. The photograph shows the pulse in two successive positions. The arrows in the diagram indicate how the coil moves as the pulse passes. The large, open arrow shows the direction of the motion of the pulse.

Fig. 6 – 4. Here we use a schematic pulse because it is a little easier to work with and we can make the time interval between positions as short as we wish. In this way we can determine the instantaneous motion of the coil. Thus, if we know in which direction the pulse is moving, we can determine how each point of the spring moves at any particular stage in the passage of the pulse. On the other hand, if we know how the parts of the spring move, we can determine the direction in which the pulse is traveling.

We now have a good notion of how the pieces of spring move, even though there is no visible motion in any one of our pictures. What we have done is to observe (1) that any pulse moves undistorted at constant speed along the spring and (2) that the spring itself moves only at right angles to the motion of the pulse. We can combine these two pieces of information to learn how each part of the spring moves at any time. Of course, we have looked only at the simplest waves, and the statement we have just made may not be true of all waves. Even in the cases we have examined, a sharp eye may detect slight deviations from our description. Nevertheless we have formed a useful first picture. With slight changes it applies to many other waves.



6–4 The relation between the motion of a pulse traveling from left to right and the motion of the spring.

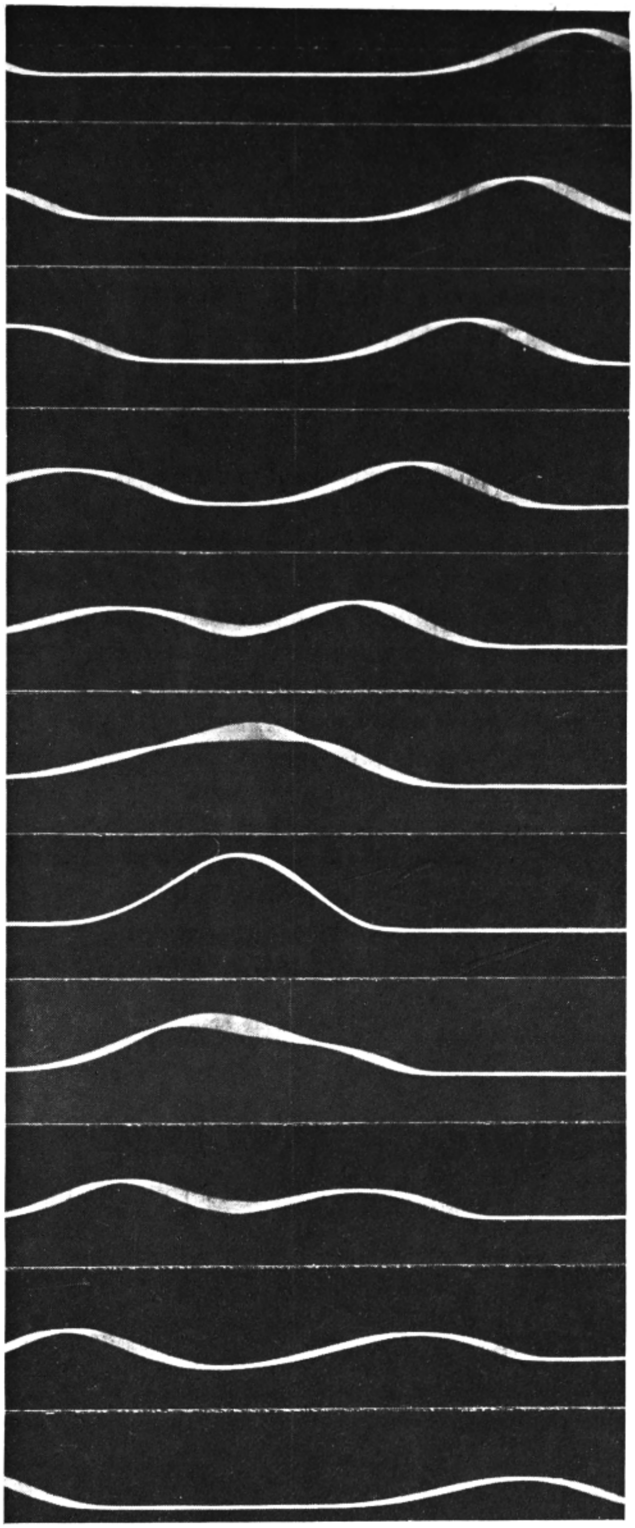
6–3 SUPERPOSITION: PULSES CROSSING

So far we have discussed the behavior of a single pulse traveling in one direction. But what happens when one pulse moves from right to left at the same time that another moves from left to right? Particularly, what happens when the two pulses meet? Do they pass through each other, or do they somehow knock each other out?

The best way to find out is to try. The photographs in Fig. 6 – 5 show what happens when two pulses are started at opposite ends of a spring at the same time, one traveling from left to right and one from right to left. The top pictures show the pulses approaching each other as if each had the spring to itself. As they cross each other, the two pulses combine to form complicated shapes. But after having crossed, they again assume their original shapes and travel along the spring as if nothing had happened, as is indicated by the pictures at the bottom. The left-going pulse continues to travel to the left with its original shape. The right-going pulse continues to move on to the right with its earlier form. We can perform this experiment over and over with different pulses and always get the same general result.

The fact that two pulses pass through each other without either being altered is a fundamental property of waves. If we throw two balls in opposite directions, and they hit each other, their motion is violently changed. The crossing of waves and the crossing of streams of balls made of solid matter are thus two very different processes.

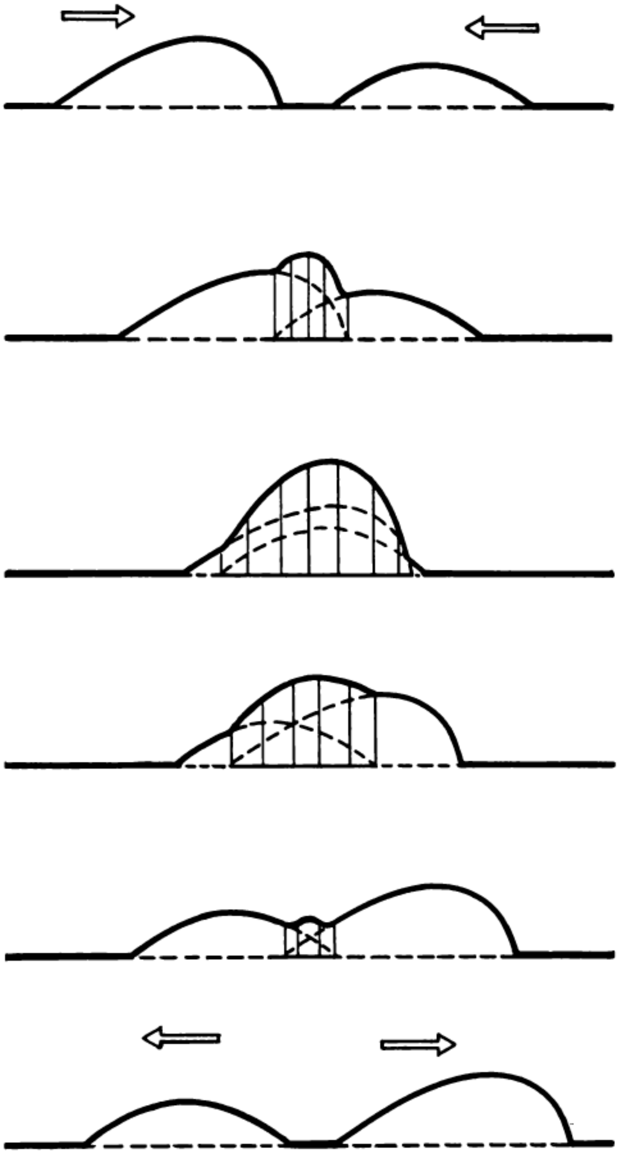
Let us now take a closer look at the two pulses crossing each other (Fig. 6 – 5). Often the shape of the combined pulse does not resemble the shape of either of the original pulses. We can see its relation to them, however, if we visualize each of the original pulses at the position it would occupy if alone and then add the displacements of the original pulses to get a new pulse. We find that the total displacement of any point on the spring at any instant is exactly equal to the sum of the displacements that would have been produced by the two pulses independently. The method is illustrated in Fig. 6 – 6. It works for any two pulses. As a matter of fact, it also works for more than two pulses—the displacements due to any number of pulses can be added.



6–5 Two pulses crossing each other. Notice that the two pulses have different shapes. Thus we can see that the one which was on the left at the beginning is on the right after the crossing, and vice versa.

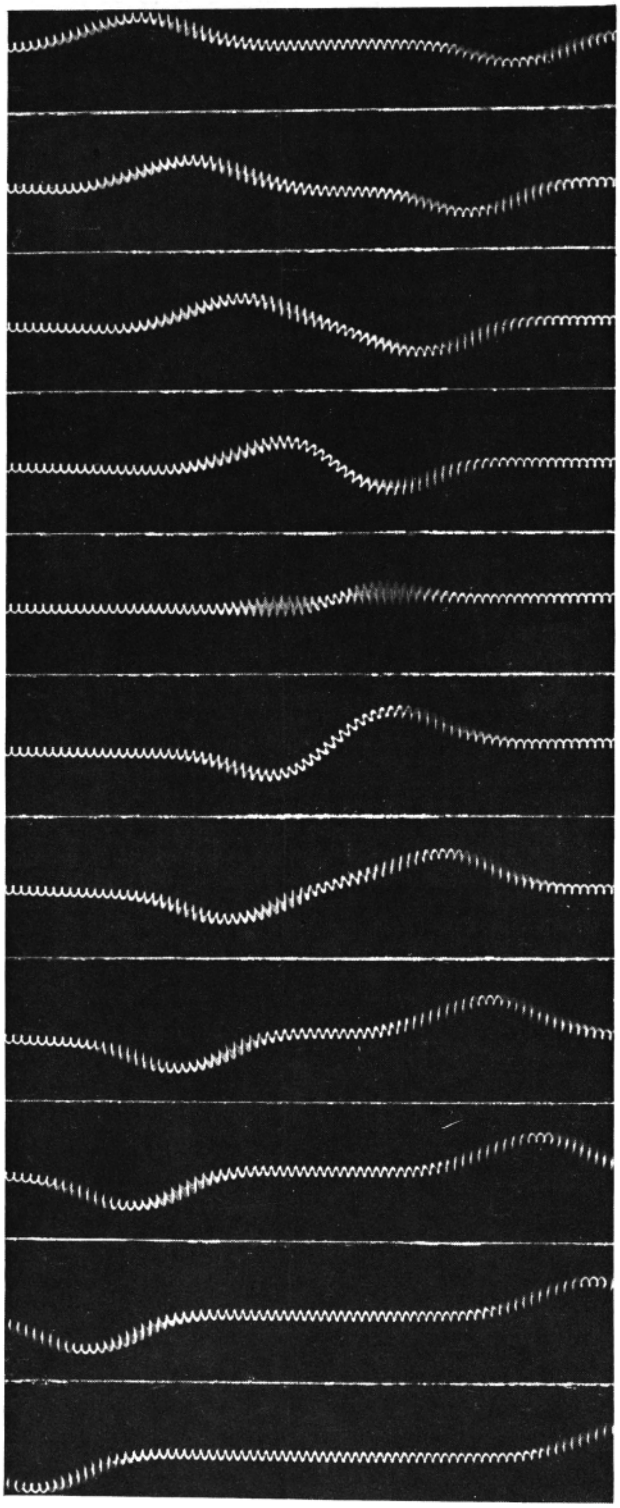
We can summarize the whole situation as follows. To find the form of the total wave disturbance at any time, we add at each point the displacements belonging to each pulse that is passing through the medium. That this simple addition gives the actual displacement of the medium is called the *Superposition Principle*.

Let us apply the Superposition Principle to two special cases. First, let us consider the



6–6 The superposition of two pulses. The displacement of the combined pulse is the sum of the separate displacements.

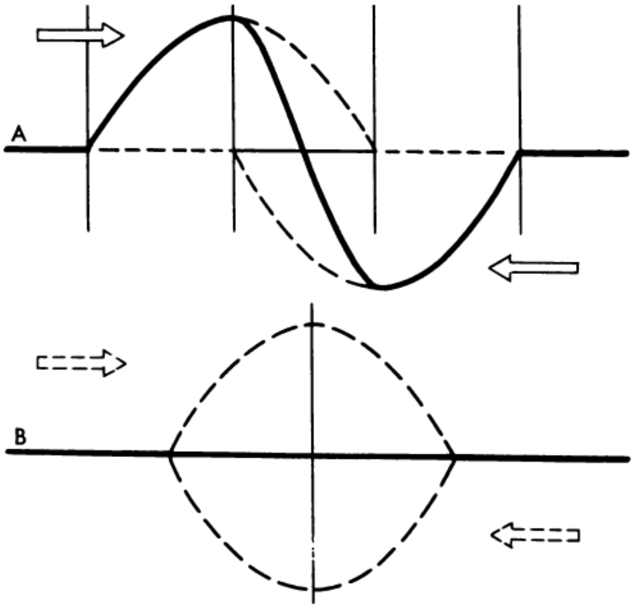
experiment shown in the sequence of pictures of Fig. 6 – 7. We consider the combination of a pulse that displaces the spring downward and travels along the spring from the right-hand end with one that displaces the spring upward and travels from the left. Suppose that the two pulses have exactly the same shape and size and that each is symmetrical. Notice that in one picture the addition of equal displacements upward (plus) and downward (minus) leaves us with a net displacement of zero. There is clearly a moment, as the pulses pass each other, when the whole spring appears undisplaced. (See also the drawing of Fig. 6 – 8.) Why does the picture not look exactly like a spring at rest? Let us consider the difference between an undisplaced spring carrying two equal and opposite wave pulses and an undisplaced spring carrying no wave at all. When the spring carries no wave, all the various pieces of spring stand still at all times. On the other hand, when two equal and opposite waves are passing, there is only one instant when the spring is passing through its rest position, and at that instant the spring is moving. The motion shows up as a blur in the pictures, just as an ordinary snapshot of a rapidly moving airplane appears blurred.



6–7 The superposition of two equal and opposite pulses on a coil spring. In the fifth picture from the top they almost cancel each other.

Our second special case is shown in Fig. 6 – 9. Here we have two similar pulses, one coming from the right and one from the left. In one the displacements are upward and in the other they are downward. These pulses differ from those in Fig. 6 – 7 in that although the two are alike in shape and size they are asymmetrical.

Because the pulses are asymmetrically shaped, they never completely cancel each

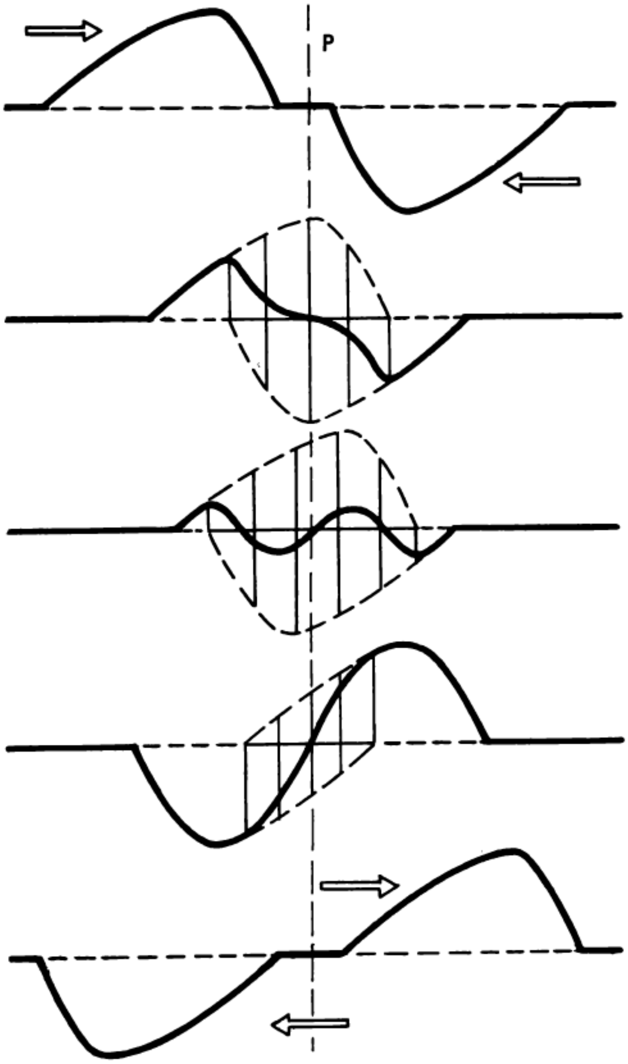


6–8 The superposition of two equal and opposite pulses. (A) Before complete cancellation. (B) At complete cancellation.

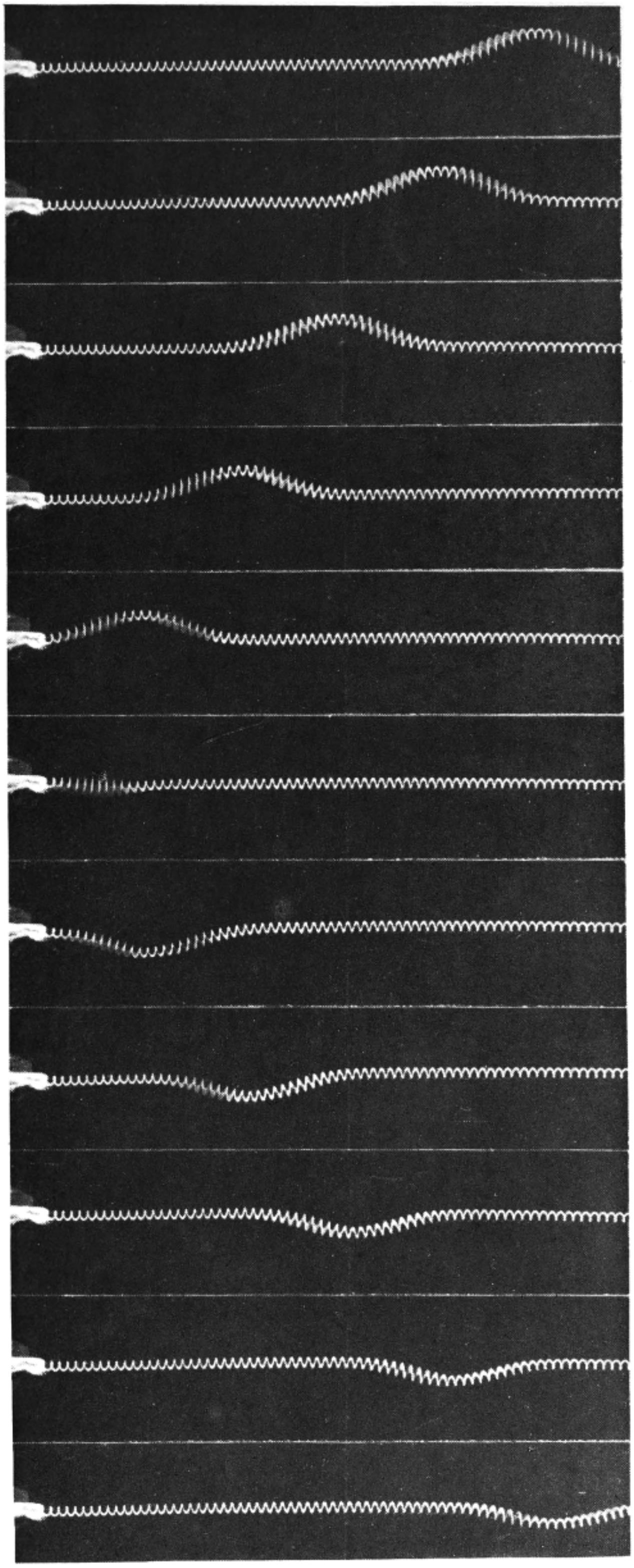
other. But there is always one point *P* on the spring which will stand still. That point is exactly halfway between the two pulses. As the pulses come together, they pass simultaneously through that halfway point in such a way that the highest point of one pulse and the lowest point of the other just cancel each other out. The same argument applies to any other pair of corresponding points on the pulses. They always arrive at the midpoint of the spring together, one on top and one at the bottom. Consequently, the midpoint stands still.

6–4 REFLECTION AND TRANSMISSION

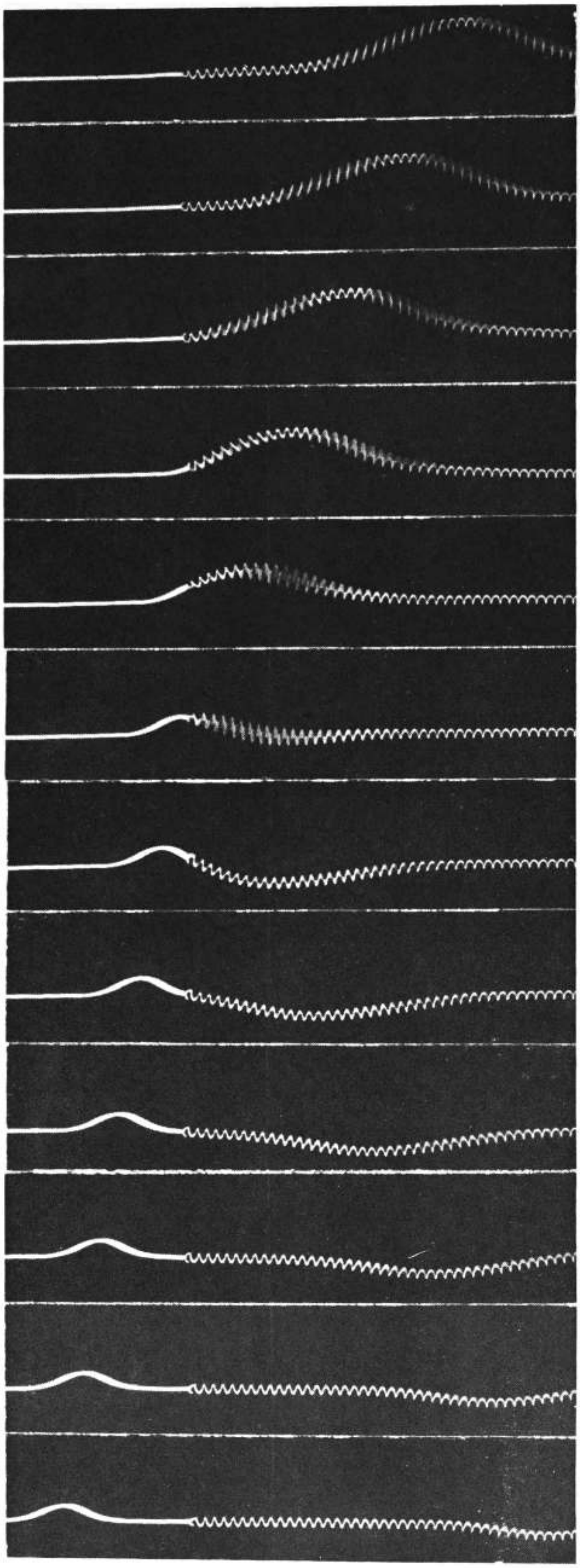
When a pulse moving on a spring comes to an end that is held fixed, it bounces back. This reversal of direction is called *reflection,* and the pulse that comes back is called the *reflected pulse*. In Fig. 6 – 10 the fixed end is on the left. In the original or *incident pulse,* which moves to



6–9 The superposition of two asymmetrical but similar pulses. Notice that the point midway between them remains at rest at all times.



6–10 Reflection of a pulse from a fixed end. The reflected pulse is upside down.



6–11 A pulse passing from a light spring (right) to a heavy spring. At the junction the pulse is partially transmitted and partially reflected. Note that the reflected pulse is upside down.

the left, the displacement is upward. The returning pulse has its displacement downward. The pulse comes back upside down, but with the same shape that it had before it was reflected.

You may wonder why the reflected pulse is upside down. The reason for this behavior is that one point on the spring, in this case the end point held by the hand, does not move. We have already encountered a situation where a point on the spring remained at rest; this was the point *P* in Fig. 6 – 9. Cover the right-hand half of Fig. 6 – 9 and you will see an upward pulse moving to the right, "flattening out" as it approaches *P*, and finally being reflected upside down. Now, at the front of an upward pulse, the spring itself moves upward (Fig. 6 – 3). When the front of the pulse in Fig. 6 – 9 gets to *P*, the point *P* should move upward. But since *P* remains at rest, the upward motion of the spring must be canceled by a downward motion. The only difference between the situations shown in Figs. 6 – 9 and 6 – 10 is that in Fig. 6 – 9 we supply the necessary downward motion by sending a downward pulse from the right, whereas in Fig. 6 – 10 we supply the downward motion by simply holding the end point fixed. Forcing the end point to remain at rest is just another way of supplying the downward motion which cancels the motion of the spring due to the original pulse, and then propagates the wave to the right in the form of an upside-down pulse.

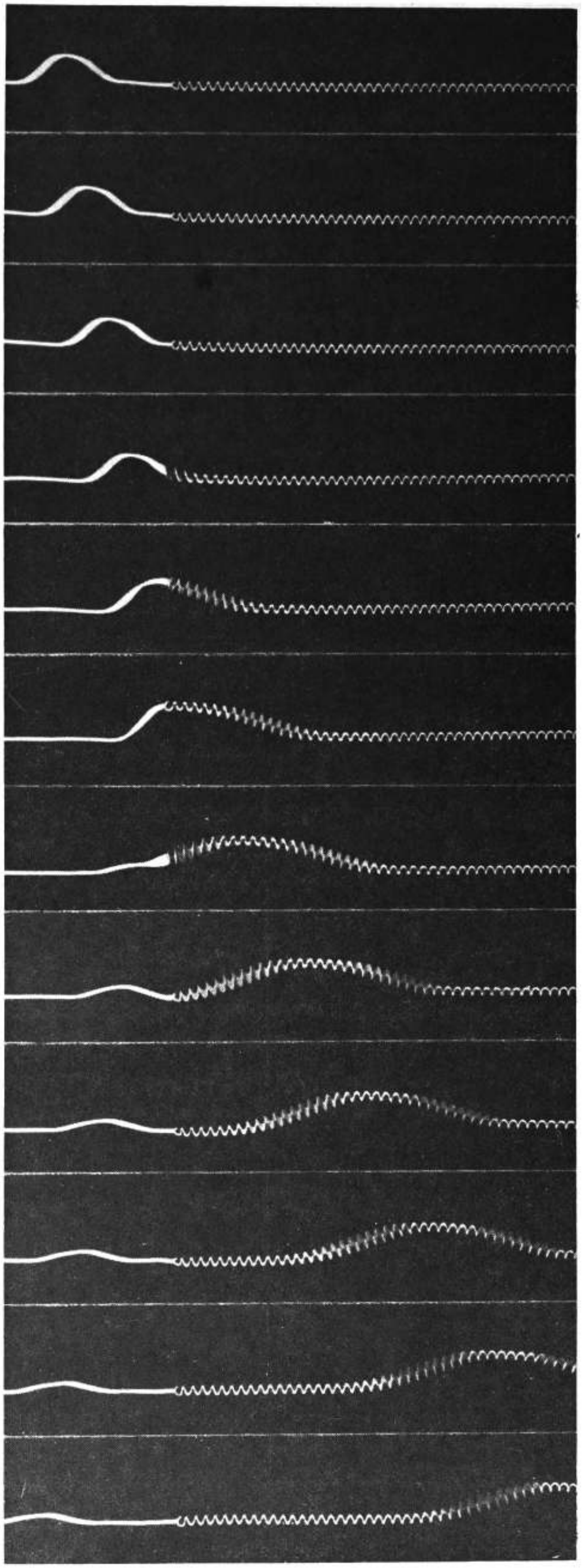
Imagine now that instead of fixing our coil spring at one end, we connect it to another spring which is much heavier and therefore harder to move. Our new arrangement will be somewhere in between the two cases (a) the original spring tied down, and (b) the original spring just lengthened by an additional piece of the same material. In case (a) the whole pulse is reflected upside down; in case (b) the whole pulse goes straight on. We may, therefore, expect that under our new arrangement part of the pulse will be *reflected* upside down, and part of it will go on, or as we say, will be *transmitted*. This effect is shown in Fig. 6 – 11 where the original pulse comes from the right and the heavier spring is on the left. We see that at the junction or boundary between the two springs – which are the media in which the wave travels – the pulse splits into two parts, a reflected and a transmitted pulse. Like superposition, the splitting into a reflected and a transmitted part is a typical wave property.

What happens when a pulse goes the other way, traveling along the heavier spring and arriving at the junction between it and the light spring? This is not so easy to foresee. We no longer can bracket the behavior between two situations in which we know the answer. But experiment tells us what takes place. In Fig. 6 – 12 we see a pulse moving from the left, from a heavy toward a light spring. Here, as in the opposite case, illustrated in Fig. 6 – 11, part of the pulse is transmitted and part is reflected, but this time the reflected pulse is right-side up.

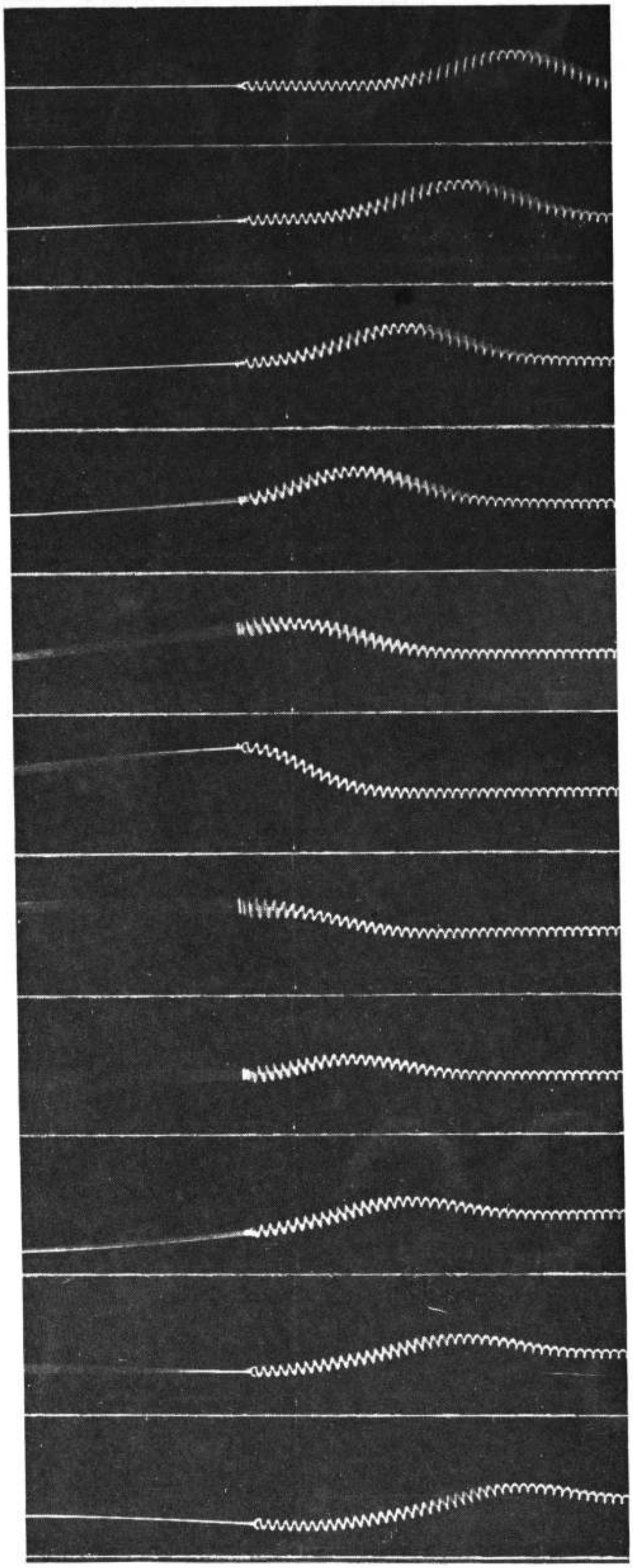
In summary, then, when a pulse is sent along a spring toward a junction with a second spring, we observe that the whole pulse is reflected upside down whenever the second spring is very much heavier than the first. As the second spring is replaced by lighter and lighter springs, the reflected pulse becomes small and a larger and larger transmitted pulse is observed to go on beyond the junction. When the second spring is only as massive as the first, no reflected pulse is left and the original pulse is completely transmitted. Then if the second spring is made still lighter, reflection sets in again, this time with the reflected pulse right-side up. The lighter the second spring, the larger is the reflected pulse. When the second spring is negligible in mass, the reflected pulse is nearly the same size as the pulse sent in. This can be demonstrated with a heavy spring tied to a thin nylon thread (Fig. 6 – 13).

6–5 IDEALIZATIONS AND APPROXIMATIONS

In discussing waves along a spring, we said that their shape or size remains unchanged during the motion. Indeed, if we look again at Fig. 6 – 1, we can notice hardly any change in the size of the pulse as it travels along. Yet, as you have undoubtedly noticed, a pulse slowly diminishes and after several reflections it dies out completely. Is it reasonable for us to ignore the dying down of the pulse? Or has our description of wave behavior been wrong in some fundamental way?



6–12 A pulse passing from a heavy spring (left) to a light spring. At the junction the pulse is partially transmitted and partially reflected. The reflected pulse is right side up.



6–13 A pulse on a spring reflected from a junction with a very light thread. The whole pulse returns right side up. The blurring of the thread in the middle frames of the sequence of pictures indicates that the particles of the thread are moving at high speed as the pulse passes. Can you determine the direction of this motion in each of the frames?

To answer these questions, we start by observing that the time it takes a pulse to die out varies with circumstances. For example, if the spring is submerged in water, the pulse dies out more rapidly than it does in air. The water offers greater resistance to the motion of the spring than the air does. We may expect that in a vacuum the pulse will require a longer time to disappear than in air; experiments to test this, although not easy to perform, justify our expectation.

Even in a vacuum, the pulse eventually dies out because of internal resistance in the spring. The amount of this resistance depends on the material of the spring. For some materials it is very small and the pulse will keep moving for a long time.

We may imagine a spring with no internal resistance kept in vacuum. On such a spring a pulse would travel forever. By ignoring the dying out of the pulses, we have idealized our real springs and considered them as if they were free from both external and internal resistance. We may do this so long as we consider the behavior of a pulse only for periods of time in which the size of the pulse changes so little that we hardly notice the change. For such times, the ideal resistance-free spring serves as a good *approximation* to the real one and can therefore be made the subject of our discussion. There is a clear advantage in this, since the behavior of the ideal spring is much simpler than that of the real one.

A similar idealization occurs in our discussion of the superposition of two pulses. We learned that the displacement produced by the combined pulse equals the sum of the displacements due to the separate pulses. But if we make the individual pulses too big, we find that the combined displacement is less than the sum of the two displacements. Again, when we ignore this deviation from the simple superposition, we are discussing an ideal spring instead of the real one. We are making an approximation rather than a complete description of the real situation. But as long as we keep our displacements small enough so that we hardly notice these deviations, the ideal spring will be a good approximation to the real one; and it has the advantage of simplicity.

Idealizations and approximations are very frequently made unconsciously. Consider, for example, what we mean when we say that the area of a piece of land is 1000 acres. Usually we get the area by measuring length and width and then calculating the area *as if* the land were completely flat. That is, we ignore the fact that there are little hills and valleys and that the area under consideration is really part of the surface of a sphere. We replace a surface of complicated shape by a simple plane rectangle. This procedure is useful only as long as there are no big mountains and the dimensions of the land are small compared with the radius of the earth. Under those conditions an idealized flat land serves as a good approximation.

Most of the problems we attack in science are fairly complicated, and in order to make progress toward understanding them we have to separate the essential from the inessential; that is, we have to make idealizations. In this chapter we have been studying waves. This is a very complicated matter on a real spring, but by mentally replacing the real spring by the idealized one, we separate the essential from the inessential and simplify the problem to help our understanding. To make the right idealization is one of the secrets of the successful scientist.

6–6 A WAVE MODEL FOR LIGHT?

In this chapter we learned about two important properties of waves which clearly indicate the advantages of a wave model of light over the particle model. First, we found that waves can pass through one another undisturbed. If we shine two flashlight beams across each other, each proceeds after the point of crossing as if the other had not been there. (See Fig. 3 – 7.) Similarly, we can see this page despite the light crossing in all directions between us and it. This means that the crossing of light beams resembles the crossing of waves much more than it does the intersection of streams of particles.

The second important wave property is that of partial reflection and partial transmission at a boundary. Recall what happens when light passes from one medium to another—say from air to glass. Part of the light is reflected and part of it is transmitted, as was shown by Fig. 4 – 11. This is just what waves do, but streams of particles do not split up this way.

These two wave properties which appear in light lead us to go on exploring a wave model for light; but they are far from demonstrating that a wave picture is an adequate model for light. For example, when a light beam hits a glass surface the angle of reflection equals the angle of incidence, and the direction of the refracted beam is described by Snell's law. On the basis of what we have studied so far, we cannot say whether a wave model accounts for the observed changes of direction. The waves on our spring are confined to move along one line or one dimension. Therefore, there is no way of changing the direction of propagation except to reverse it. To find if waves can really account for the behavior of light, we must have waves which move in space or at least in a plane, so that we can make a direct comparison. This we shall do in the next chapter, where we shall study waves on the surface of water.

QUESTIONS

1.* What is similar in all examples of waves discussed in Section 1?

2. Suppose you look out your window and see your neighbor across the street sitting on his porch. In how many ways could you do something to attract his attention, make him move, or otherwise influence his actions? Which ways involve mass transmission and which ways wave motion?

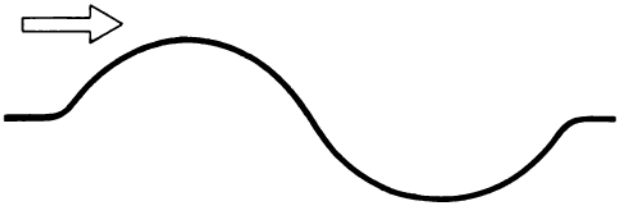
3.* According to Fig. 6 – 4, along what part of the pulse is the spring momentarily at rest?

4.* In Fig. 6 – 2, the ribbon first moves up and then down as the wave moves past it from right to left. Would the ribbon move up first or down first if the same pulse traveled from left to right?

5. Figure 6 – 2 shows the displacement of a point on a spring as a pulse goes by. Make a graph showing the displacement of this point as a function of time. Plot displacement vertically and time horizontally with $\frac{1}{24}$ -sec intervals (the interval between pictures of Fig. 6 – 2).

6. Sketch the motion of the spring for the pulse in Fig. 6 – 14.

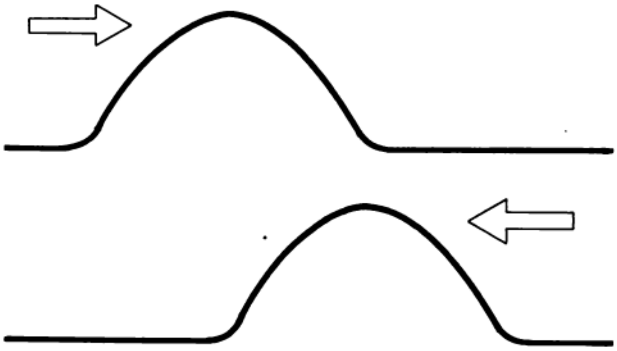
7.* If two pulses traveling toward each other on a coil have displacements in the same direction, can they cancel each other when they cross?



6–14 **For Question 6. A pulse moving to the right.**

8.* Two pulses have maximum displacements of 3 cm and 4 cm in the same direction. What will be the maximum displacement when they pass?

9. Using the two pulses shown in Fig. 6 – 15, determine the size and shape of the combined pulse at this moment. Do the same thing for several other positions of the pulses.

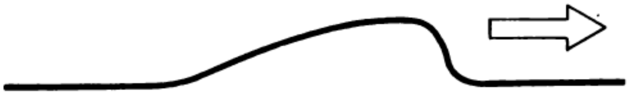


6–15 **For Question 9. Two equal pulses moving in opposite directions.**

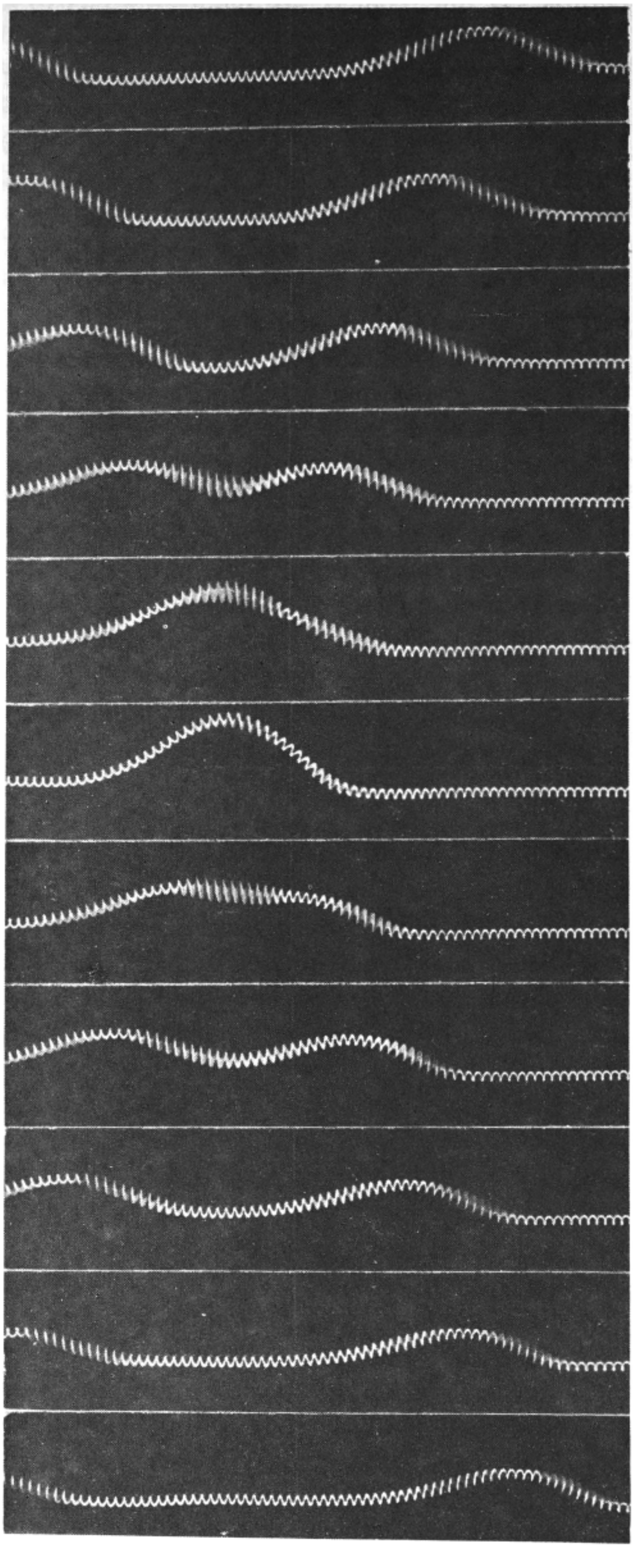
10. The seventh picture from the top in Fig. 6 – 5 shows two pulses at the moment of crossing. Specify the pieces of spring that are moving and their direction of motion.

11. In the fifth picture from the top of Fig. 6 – 7, which points are moving and in which direction do they move?

12.* A pulse, shown in Fig. 6 – 16, is sent along a coil toward the right. Draw the pulse traveling to the left which could momentarily cancel the pulse shown.



6–16 **For Question 12.**



6–17 **For Question 13. Superposition of two equal and symmetric pulses.**

13. In the sixth picture from the top, Fig. 6 – 17, we see the superposition of two equal pulses, each of which is symmetrical about its center line.

    (a) The absence of blur indicates that there is no motion at this instant. Show that this is true by using the principle of superposition.

    (b) Assume that you deform the coil spring in the same manner as shown in the sixth picture. What will happen when it is released?

14. The sixth picture of Fig. 6 – 10 shows the spring at an instant when the spring is almost straight. Explain why there is an instant when this happens.

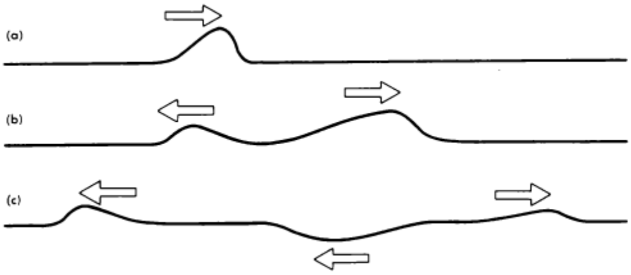
15. Consider the asymmetric pulse coming from the left in Fig. 6 – 6. Draw the shape it will have after being reflected at a fixed end.

16.* In Fig. 6 – 11, which has the larger displacement, the incident or the transmitted pulse?

17.* In Fig. 6 – 11, what is the ratio of the speed of the pulse in the light spring to the speed in the heavy spring?

18. You send out a pulse at one end of a coil, and it returns to you upside down and smaller in size. What can you deduce about the speed of the pulse on a second coil which is attached to the other end of your coil?

19. Diagram (a) in Fig. 6 – 18 shows a pulse moving along a rope which has sections of different densities. Diagrams (b) and (c) show the same rope at equal intervals of time later. Where are the junctions, and what are the relative densities of the rope between them?



6–18 For Question 19.

20. When light passes from air to water or vice versa, part of it is reflected. If this situation resembles that of a pulse crossing from one coil spring to another, in which case will you expect the light pulse to be reflected upside down?

21. Hold one end of a long rope with the other end tied to a rigid support. Stand looking along the rope and generate a wave by moving your hand through three fast clockwise circles.

    (a) Describe the wave generated.

    (b) Describe the reflected wave.

    (c) Describe the motion of a particle of the rope as the wave passes forward and back.

22.* How could you tell in Fig. 6 – 10 that the sequence of events proceeds from top to bottom and not from bottom to top?

23. Whenever a pulse travels along a spring, its size decreases as it moves along.

    (a) Under what conditions are we justified in neglecting this decrease?

    (b) What is the advantage of neglecting this decrease?

24. What do we mean by an ideal spring in the context of this chapter?

25. We can say that the surface of the sea is approximately flat. Give some examples where this is a good approximation and also some where it is a bad approximation.

7

Waves and Light

7–1 WATER WAVES

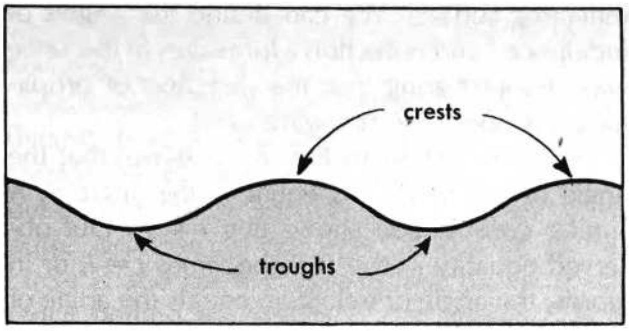
One of the first illustrations of waves mentioned in Chapter 6 was that of ripples on a pond. The spreading out of waves in the form of larger and larger circles is familiar to everyone. For example, a fish nibbling at a worm on a hook gives away its presence to the angler who sees the circular waves produced by the up-and-down motion of the float attached to the line.

Because water waves move along the surface and do not extend downward to any appreciable depth, they are known as surface waves. If you have watched fish in an aquarium, perhaps you have noticed that they are undisturbed by the waves. A submarine commander does not fear a stormy sea as the captain of a surface ship might. His submarine can dive and travel along unaffected by the powerful waves above.

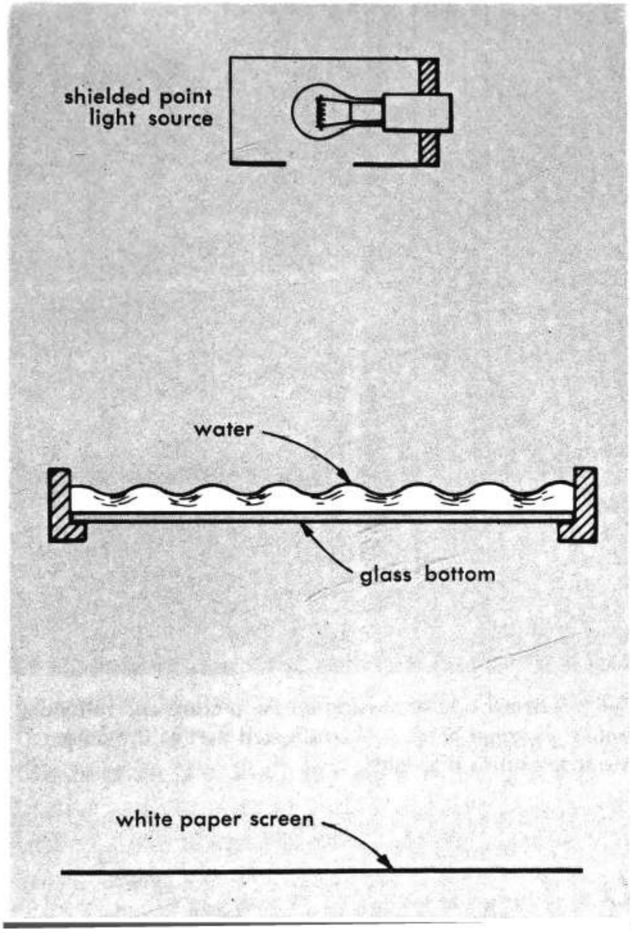
If we look at water waves through the side of an aquarium, we are able to see their shape. We notice that, while there are some variations, they are generally similar and look something like the illustration in Fig. 7 – 1. The upper portions of the waves are called *crests* and the lower portions *troughs*. To produce and observe waves in the laboratory we use an apparatus called a *ripple tank*. The ripple tank (Fig. 7 – 2) has a glass bottom which makes it possible to project images of the waves onto a screen. These images are produced because the crests of the waves act as converging lenses and tend to focus the light from the lamp while the troughs, acting as diverging lenses, tend to spread it out. Therefore the crests appear on the screen as bright bands while the troughs appear dark.

7–2 STRAIGHT AND CIRCULAR PULSES

We can generate a straight pulse, like the long bow waves caused by a passing boat, by dipping a ruler horizontally into the surface of the water of the ripple tank. The motion of the pulse is such that its crest always


crests
troughs

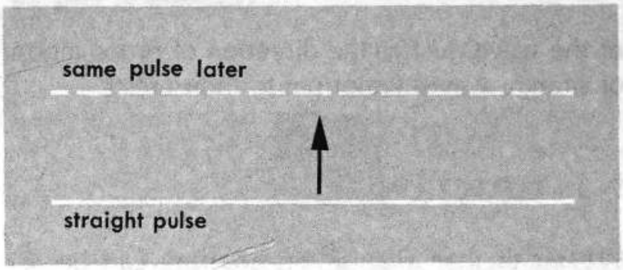
7–1 An illustration of water waves.


shielded point light source
water
glass bottom
white paper screen

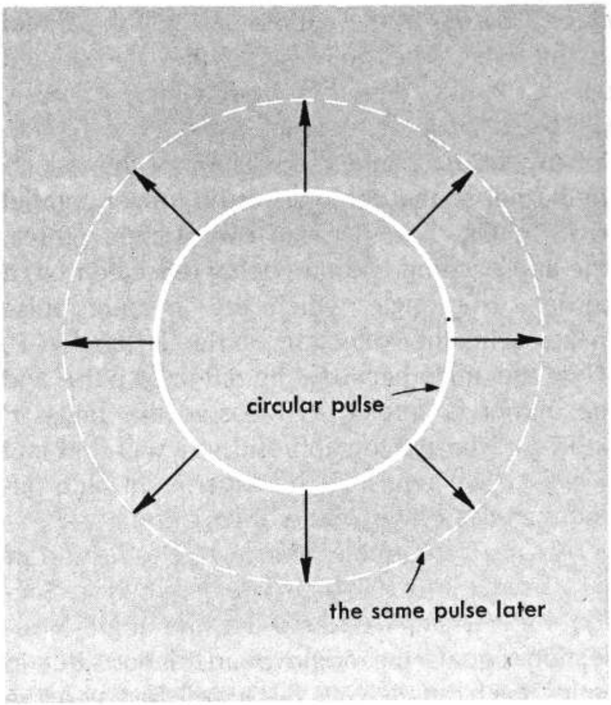
7–2 The ripple tank.

remains parallel to a line marking its original position. The distance between parallel lines is measured along a perpendicular; thus the direction of motion of the pulse (also called the *direction of propagation*) is perpendicular to the wave crest. This direction is called the *normal* to the crest. In Fig. 7 – 3 the crest of a straight pulse is shown as a solid white line. Its position at a later time is shown by the dashed line. The direction of propagation is shown by the arrow. Whenever we draw a straight pulse we shall indicate its direction of propagation by a small arrow.

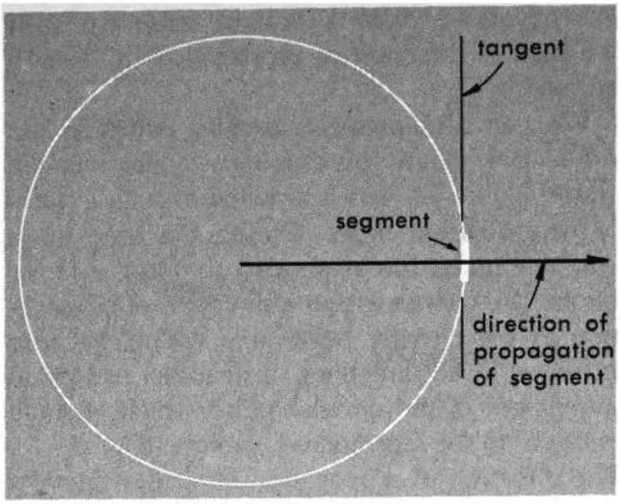
We can also produce circular pulses in the ripple tank simply by dipping a finger into the water. Figure 7 – 4 is a drawing of such a pulse at two different times. During the time interval, the pulse has expanded to form a larger circle. We cannot assign a direction of propagation to the circular pulse as a whole, because it moves in all directions. Instead, let us look at a segment of the circular pulse which is small enough to be considered straight (Fig. 7 – 5). The direction of propagation of such a segment is along the radius and away from the center of the circle. This direction is normal to the crest


same pulse later
straight pulse

7–3 A straight pulse moves at right angles to its crest. The arrow indicates the direction.


circular pulse
the same pulse later

7–4 An expanding circular pulse.



7–5 A small segment of a circular pulse. The segment acts as though it were straight and moves along a radius away from the center.

of the wave, just as the direction of propagation of straight waves is normal to their crests.

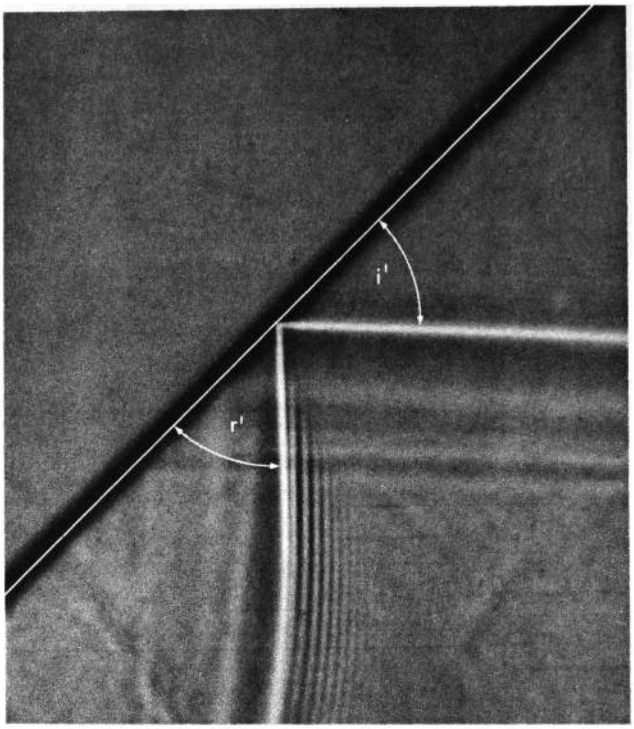
7–3 REFLECTION

We recall that a pulse on a spring can be reflected, and we may expect that water waves can also undergo reflection. Consider a straight pulse as it moves away from the ruler toward the opposite end of the tank. To reflect it we place a barrier in the middle of the tank parallel to the ruler. The pulse strikes the barrier and reflects in the direction from which it came, just like a pulse on a spring.

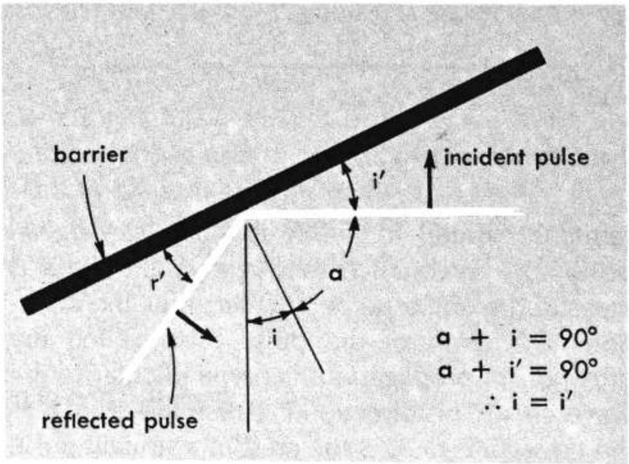
Now let us change the position of the reflecting barrier so that the pulse is no longer parallel to it. In Fig. 7 – 6 we see two straight pulses, one approaching and one being reflected from a barrier. The angle which the incident pulse makes with the reflecting barrier is labeled i', while the angle between the reflected pulse and the barrier is labeled r'. Measure the angles r' and i' on the photograph and you will find that $r' = i'$. If we repeated the experiment with different angles i', we would always find $r' = i'$.

This result resembles the law of reflection of light from mirrors, which we obtained in Section 4 – 1. There we found that the angle of reflection equals the angle of incidence. But in optics each angle was measured between the direction of propagation and the normal to the reflecting surface. We can define the angles of incidence i and reflection r for waves in the same way, remembering that the direction of propagation is normal to the wave crest.

The construction in Fig. 7 – 7 shows that the angle of incidence i is equal to the angle i'. A similar construction shows that $r = r'$. Our observed equality $r' = i'$ then becomes $r = i$, or in words, the angle of reflection equals the angle of incidence. We have shown, therefore, that



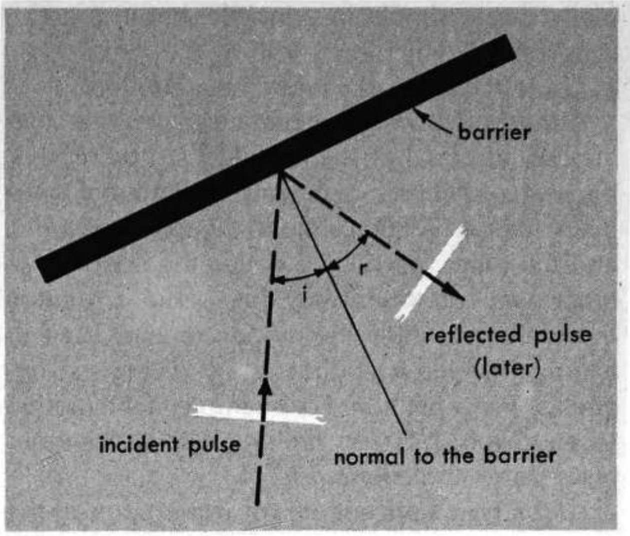
7–6 A straight pulse moving up the picture and reflecting from a diagonal barrier. The reflected part of the pulse is moving off to the right.



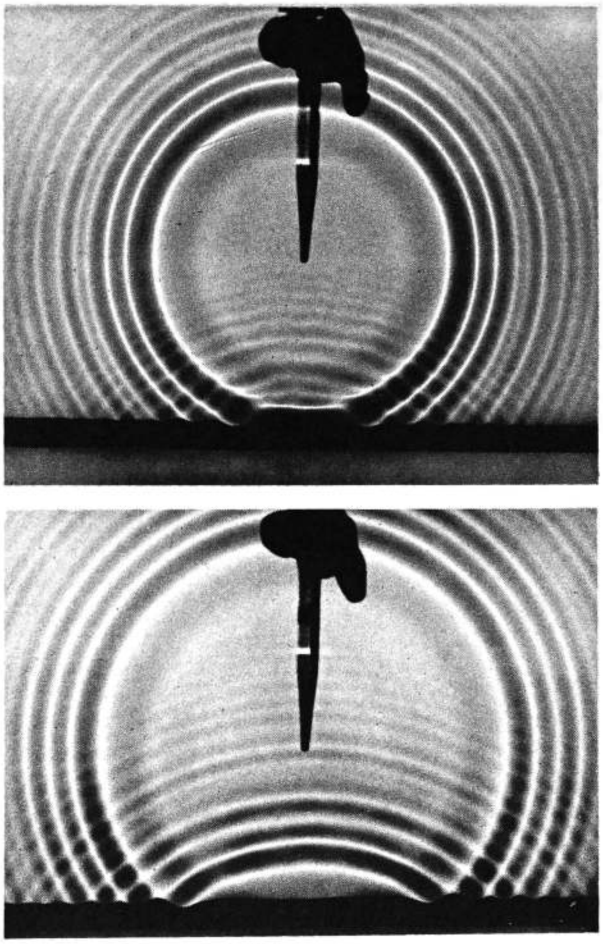
7–7 A straight pulse incident on a barrier. The angle of incidence i is equal to the angle the pulse makes with the barrier i'.

waves and light follow the same law of reflection (Fig. 7 – 8).

An expanding circular pulse can also be reflected from a straight barrier. Figure 7 – 9 shows the approach and reflection of such a pulse. We notice that the reflected part of the pulse is an arc of a circle. The center of this circle is at a point *P′*, which is at the same distance behind the barrier as the source *P* is in front of it. The reflected pulse appears to come from *P′*. This corresponds to the situation in optics where we placed a point source in front of a mirror. The reflected light then appeared to come from the image point behind the mirror. (See Figs. 4 – 3 and 4 – 4.) By using curved barriers or combinations of two or more straight barriers, we can demonstrate with the ripple tank all of the phenomena of reflection that we studied in connection with light. Just as the formation of images by mirrors followed from the laws of reflection in optics, so does the corresponding formation of images by "mirrors" in the ripple tank.



7–8 The reflection of a straight wave from a straight barrier. Just as with light, the angle of reflection r equals the angle of incidence i.



7–9 Reflection of a circular pulse from a straight barrier. In the upper photograph the strongest pulse has just struck the barrier, while in the lower one a part of it has been reflected.

7–4 SPEED OF PROPAGATION AND PERIODIC WAVES

Waves in different media propagate with different speeds. For example, we can see the waves on a coil spring speed up when we stretch the coil, and we can see the waves in a rubber hose slow down when we fill it with water. In this section we shall learn how to make quantitative measurements of the speed of water waves in the ripple tank. There are several ways of going about such a measurement.

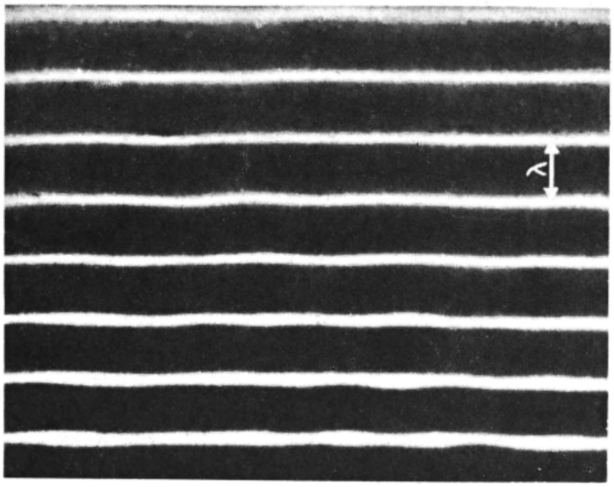
One way is to generate a straight pulse and measure with a stopwatch the time t the pulse takes to travel a specified distance l. The speed v is then equal to the distance traveled divided by the time taken

$$v = l/t$$

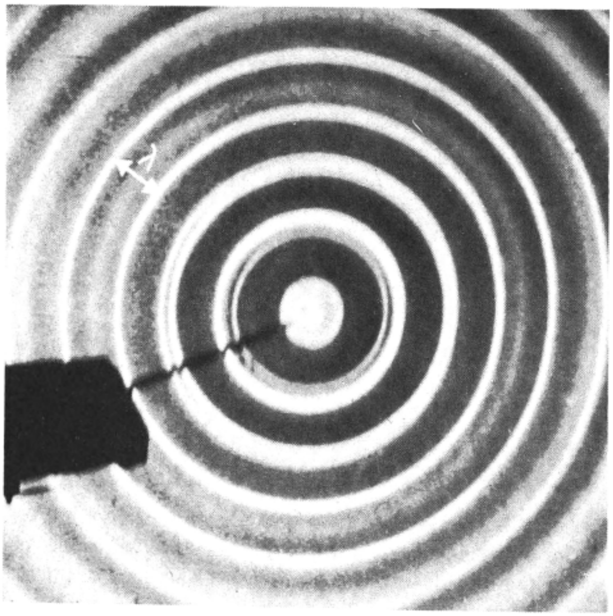
Another way is to generate two straight pulses, one after the other. By the time the second pulse is generated (after a time t) the first pulse has traveled a distance l. From then on both pulses travel along, while the distance l between them remains the same. We can measure this distance with a ruler, and again $v = l/t$. These methods are simple in principle, but in practice it is rather difficult to follow the pulses and measure the distances and times required.

A third method is to generate pulses one after another at equal time intervals T. In doing this, the wave generator repeats its motion once every interval T. Such a motion is called *periodic,* and the time interval T is called the *period.* Another way of describing this periodic motion is to tell how often the motion repeats itself in a unit time interval; that is, by giving the *frequency* f of repetition. For example, if the motion repeats every 1/10 sec, the frequency is ten times per second. In general $f = 1/T$.

Let us now concentrate on some point in the tank. The pulses produced by the generator move toward this point, and they pass the point with the same frequency with which they leave the source. If ten are sent out each second, ten will pass each second. The frequency of the wave is therefore also given by $f = 1/T$, and T is the time between the passage of successive waves. Furthermore, as the waves move, the distance between any two adjacent pulses is always the same and is called the *wavelength* λ (lambda). The wave pattern which we have been describing is called a periodic straight wave (Fig. 7 – 10).



7–10 Periodic straight waves moving across a ripple tank.



7–11 Periodic circular waves.

We can obtain the speed of a periodic wave in a manner similar to that which we used for a pair of pulses. We know that the pulses are separated by a distance λ and that each pulse moves over this distance in a time T. Hence the speed of propagation is

$$v = \lambda/T$$

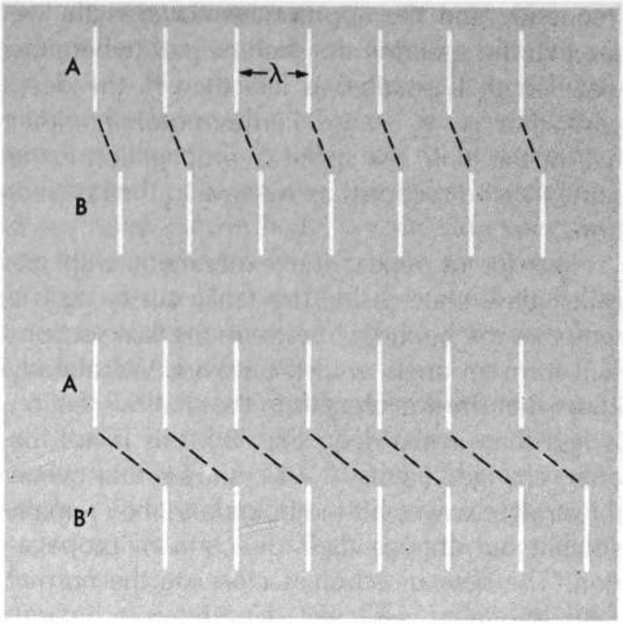
Using the relation $f = 1/T$, we find that

$$v = f\lambda$$

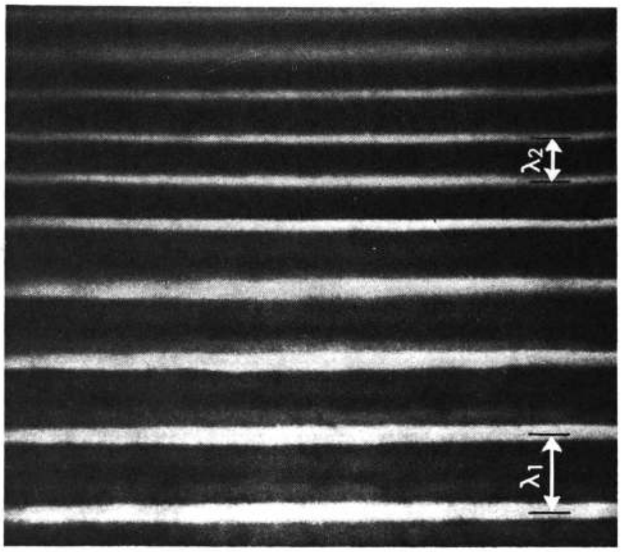
or that the speed of propagation of a periodic wave is the product of the frequency and the wavelength.

The relation that we have just obtained is by no means restricted to waves in a ripple tank. It is equally good for *any* periodic wave. Such things as the straightness of the wave, the nature of the ripple tank, and the properties of the water did not enter into the argument from which we got our result. In particular, we could have followed the same procedure with circular periodic waves and would have found the relationship $v = f\lambda$ again. In this case the wavelength is measured along the radius (Fig. 7 – 11); and we find it is equal to the wavelength of a straight wave of the same frequency. The speed of circular waves is therefore equal to that of straight waves in the same medium. Furthermore, we could have applied the above arguments to any other kind of periodic waves—for example, periodic waves on coils—and again would have found the relation $v = f\lambda$.

Now we come to the advantage of the above relation for the measurement of v. Imagine that instead of watching the wave continuously we look at it through a shutter which is closed most of the time and opens periodically for short time intervals. The stroboscope described in Chapter 1 is such a device. The first time the shutter opens we will get a glimpse of the wave pattern in a certain position. During the time the shutter is closed, all of the pulses will move a distance equal to their speed times the time the shutter is closed (Fig. 7 – 12). If we look through the shutter while it is periodically opening and closing, the pattern will usually appear to move. Suppose, however, that the period of the shutter is just the same as that of the wave. Then, during the time the shutter is closed, every pulse just moves up to the position of the pulse ahead of it, and we see the same pattern every time the shutter opens. That is, we see a stationary pattern from which it is easy to measure the wavelength. Moreover, as we have said, the period of the shutter is equal to the period of the wave and it can be measured by counting the number of times the shutter is opened in a given time interval; i.e., by measuring the frequency of the shutter. This gives us *f*. Now that we have both λ and *f* for the wave, we can make use of the general relation $v = f\lambda$ in order to determine the speed.



7–12 Crests of a periodic wave seen at successive openings A and B of a stroboscope shutter. In the top diagram the frequency of the stroboscope is greater than that of the waves. At the bottom, it is the same. The dashed lines, of course, are not visible.



7–13 Passage of waves from deep to shallow water. The deep water is at the bottom and the shallow water at the top of the picture. Note that the wavelength is shorter in the shallow water.

7–5 REFRACTION

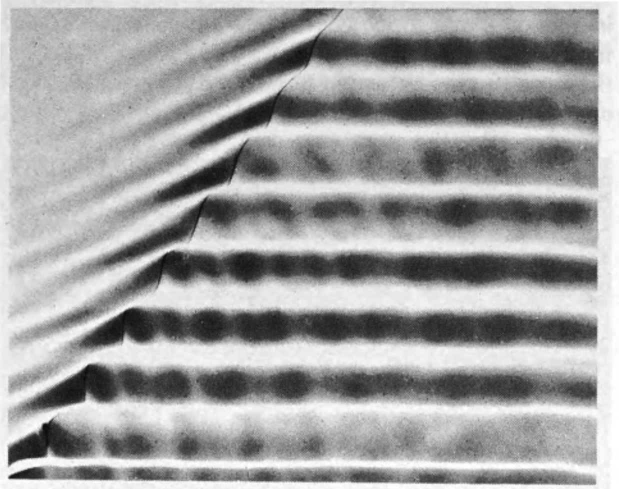
We mentioned earlier that the speed of propagation of waves depends on the properties of the medium through which they move. In the case of waves on the surface of water, the speed depends on the depth of the water. Therefore water of two different depths can be considered to be two different media for wave propagation. This is a very useful property, because by merely changing the depth of the water in part of the ripple tank we are able to study the behavior of waves when they pass from one medium to another.

To see that the speed of propagation indeed depends on the depth, we now make half the tank shallower by placing a thick glass plate on the bottom of the back part of the tank, thereby dividing it into two sections of different depths with the dividing line parallel to the waves. Let us look at the waves through the stroboscope. By turning it at the right frequency we can stop the motion of the pattern in *both* sections of the tank simultaneously. Hence the frequency is the same in both sections; it is not affected by the change in depth. But Fig. 7 – 13 shows that the wavelength λ_2 in the shallow part is shorter than λ_1, the wavelength in the deep part.* Since the speed of propagation is the product of the same frequency and the appropriate wavelength, we see that the speed in the shallow part (where the wavelength is smaller) is less than in the deep part—that is, $v_2 < v_1$. Furthermore, in either half of the tank, the speed of propagation is the same in all directions, as we saw in the last section.

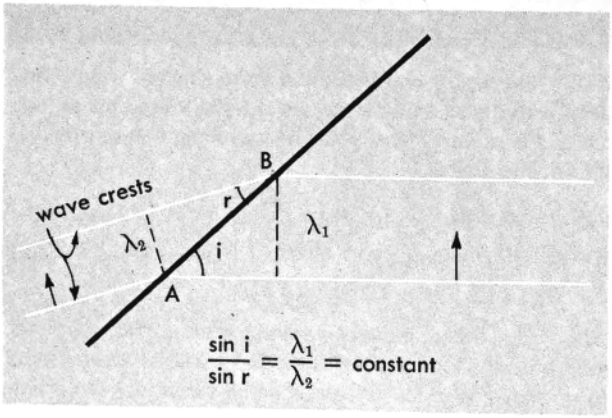
---

*You will notice that the waves tend to disappear toward the end of the shallow section. They die down only because we do not have an ideal state of affairs in which we can isolate the phenomenon we wish to study. Here, as in the experiments with the coil spring, there is a certain amount of resistance present which reduces the motion and finally causes it to fade completely.

Now let us repeat the experiment with another glass plate inside the tank, cut in such a way that the boundary between the two sections will form an angle with the waves. We already know that the wavelength in the shallow section is less than in the deep one. But this is not the only change. Figure 7 – 14 shows that when the straight waves hit the boundary, they remain straight but change their direction of propagation. The new direction is closer to the normal to the boundary than the original direction of propagation. We remember from our study of optics that this is what happens to light when it passes from one medium to another in which its speed is less (Sections 4 – 4 and 5 – 7). For light this refraction is quantitatively described by Snell's law. The following question now suggests itself. Does Snell's law also hold true for waves such as those in the ripple tank, when they pass from one medium to another?



7–14 Refraction of waves at a boundary between deep and shallow sections of the ripple tank. Note the weak reflected waves.


wave crests
B
r
λ_2
λ_1
i
A
$\frac{\sin i}{\sin r} = \frac{\lambda_1}{\lambda_2} = \text{constant}$

7–15 The geometry of the refraction of two consecutive wave fronts.

There are two procedures open to us at this point. We can measure many angles of incidence and angles of refraction in the ripple tank, and thus find experimentally whether they are related by Snell's law; or we can find out whether Snell's law can be predicted theoretically from the properties of waves which we already know. We shall take the second course here.

In Section 7 – 3, we proved that the angle of incidence is equal to the angle between the incident wave crest and the barrier. Likewise, the angle of refraction is equal to the angle between the refracted wave crest and the barrier. Let us now draw two consecutive wave fronts as they are refracted at the barrier (Fig. 7 – 15). (There is no need to add the normals, since angles equal to i and r already appear in the drawing.) Then by definition

$$\sin i = \frac{\lambda_1}{AB}, \quad \sin r = \frac{\lambda_2}{AB}$$

The values of $\sin i$ and $\sin r$ change from case to case, but their ratio is a constant independent of the angle of incidence, as we see by actually dividing $\sin i$ by $\sin r$:

$$\frac{\sin i}{\sin r} = \frac{\lambda_1\, AB}{AB\, \lambda_2} = \frac{\lambda_1}{\lambda_2} = \text{constant}$$

The relation $\sin i/\sin r = $ constant is Snell's law, this time for waves. We shall again call the constant the index of refraction and denote it as in optics by n_{12}. Thus

$$\frac{\sin i}{\sin r} = n_{12}, \text{ and } n_{12} = \frac{\lambda_1}{\lambda_2}$$

We can now express the value of n_{12} in terms of the speeds of propagation of the waves in the two media. In general the speed v is related to the wavelength as $v = f\lambda$. In particular, then, $\lambda_1 = v_1/f$, and $\lambda_2 = v_2/f$. Hence

$$n_{12} = \frac{\lambda_1}{\lambda_2} = \frac{v_1/f}{v_2/f} = \frac{v_1}{v_2}$$

This equation says that the index of refraction is equal to the ratio of the speed of propagation in the first medium to that in the second.

We arrived at Snell's law and the relation of n_{12} to the speeds by theoretical analysis of our previous results. We could get the same conclusions by direct measurements. First we could measure many pairs of angles i and r and thus establish that $\sin i/\sin r = \text{const.} = n_{12}$. Then we could measure the speeds of propagation in the two media and establish that $n_{12} = v_1/v_2$. Such measurements have often been carried out, and they do agree with our conclusions.

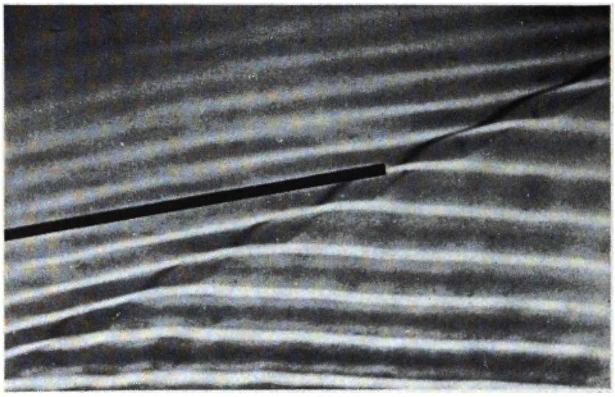
We now recall that the particle model of light as developed by Newton (1669) also explained the existence of a constant index of refraction for a given pair of substances but predicted its value to be $n_{12} = v_2/v_1$ (Sections 5 – 2 and 5 – 7). Our wave model, advocated by Huygens (1677) predicts $n_{12} = v_1/v_2$, just the reverse of Newton's result. The position of the wave model of light was strengthened by various experiments at the beginning of the nineteenth century. But this particular question was not settled for almost two hundred years. In 1862 Foucault actually measured the speed of light in air and in water and found that the speed in water was less. The exact ratio $v_1/v_2 = 1.33$ was measured by Michelson in 1883. This ratio agrees with the wave model because the index of refraction of water is 1.33 (Chapter 13).

One point still needs clarification: not all the light hitting the boundary between the two media is refracted. Part of it is reflected even if both media are transparent. The same holds for waves. In Fig. 7 – 14 the size of the reflected wave is rather small, but you can see it if you look closely. We conclude, therefore, that as far as refraction is concerned, waves have just the properties which we need in order to explain the behavior of light.

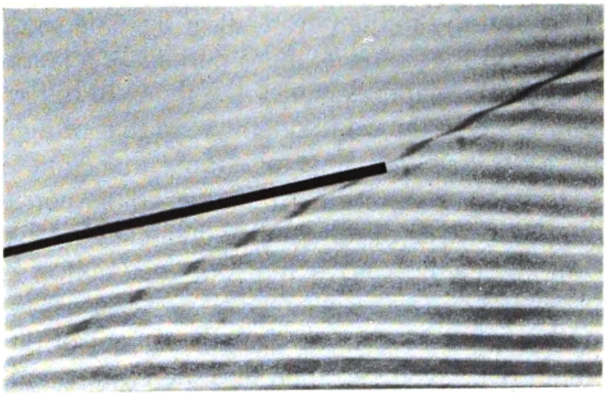
7–6 DISPERSION

In the last section we studied the refraction of periodic waves as they pass from one medium to another. We found the index of refraction to be equal to the ratio of the speeds of propagation in the two media: $n_{12} = v_1/v_2$. We did not state the frequency of the waves, because we had previously learned that the speeds of propagation depend only on the media in which the waves travel. Accordingly, we should expect to find the same index of refraction for waves of different frequency provided we repeat the experiment with the same two media, for example, water of the same two depths.

What we actually observe is shown in Figs. 7 – 16 and 7 – 17. In Fig. 7 – 16 we see the refraction of a wave of low frequency (long wavelength). To indicate the direction of the refracted waves we placed a rod on the screen of the ripple tank. It is exactly parallel to the refracted wave fronts. We then increased the frequency (i.e., decreased the wavelength), leaving the rod untouched. Notice that in Fig. 7 – 17 the rod is no longer parallel to the refracted wave crests. The wave with the higher



7–16 Refraction of low-frequency waves. The black marker is placed parallel to the refracted waves.



7–17 Refraction of high-frequency waves. The refracted waves are not parallel to the black marker.

frequency is clearly refracted in a direction slightly different from that taken by the low-frequency wave, although the angle of incidence is the same in both cases. The index of refraction for the two media therefore depends somewhat on the frequency of the wave. By analogy with the dependence of the index of refraction on the color of light, which we have discussed in Section 4 – 10, this phenomenon is called *dispersion*. Since the index of refraction equals the ratio of the speeds of propagation in the two media, we have to conclude that the speed must depend on frequency in at least one of the two media; otherwise the ratio could not show such a dependence. A medium in which the speed of waves depends on the frequency is called a *dispersive* medium.

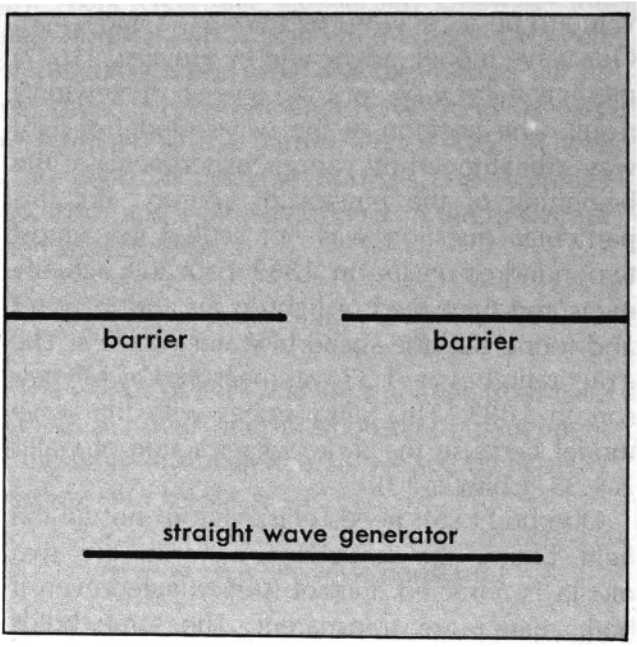
In the ripple tank we can measure the speed of periodic waves of different frequencies (Section 7 – 4), and thus see directly that the speed changes with the frequency, provided we make our measurements accurately enough. The statement we so often make, that the speed of waves depends only on the medium, is therefore an idealization. To be sure, this idealization is a good approximation to the true state of affairs provided we are not concerned with small changes in the speed (Section 6 – 5).

We made a similar idealization in our study of the refraction of light in Chapter 4. You will recall that the index of refraction of light at first appeared to be dependent only on the two media through which the light was passing—for example, air and glass. Then a closer examination of refraction, using prisms, showed that the index changes slightly with color; it is a little larger for violet than for red.

The index of refraction of waves depends slightly on the frequency. That of light depends slightly on the color. Is there perhaps a relation between the dependence of the index of refraction of waves on the frequency and that of light on color? It is tempting to assume that light is a periodic wave and that different colors correspond to waves of different frequencies. At this point we cannot prove that our assumption is correct, but in the next chapter we shall learn how to measure the wavelength of light of different colors. We can then find the corresponding frequencies, and we shall indeed establish that light waves of various frequencies appear to our eyes as light of various colors.*

7–7 DIFFRACTION

Our study of refraction and dispersion clearly shows that the wave picture of light succeeds where the corpuscular picture fails. Yet the corpuscular picture predicts correctly that light should propagate in straight lines and cast sharp shadows. Can a wave model also account for these properties of light? Again, a good way to investigate these questions is to experiment with waves in the ripple tank. We use a straight wave generator and two barriers parallel to it placed in line with an opening between them (Fig. 7 – 18). These barriers would cast sharp shadows if light were incident upon them from the direction of the generator. What happens when we send in a periodic straight wave of



7–18 Ripple-tank arrangement for the experiment shown in Fig. 7–19.

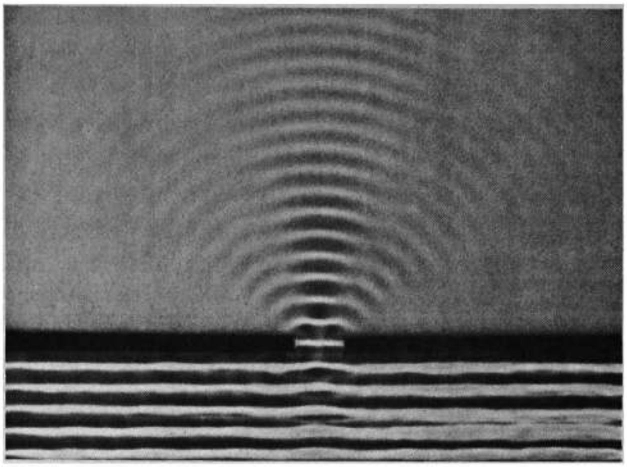
---

*By comparing the bending of waves in Figs. 7 – 16 and 7 – 17 with the dispersion of light by glass (see Fig. 4 – 14), we might come to the conclusion that the wavelength of violet light is longer than that of red light. However, blue light does not refract more than red in all materials. In some materials it is the other way around. Therefore, studying the dispersion of water waves does not establish the relation between color and wavelength of light.

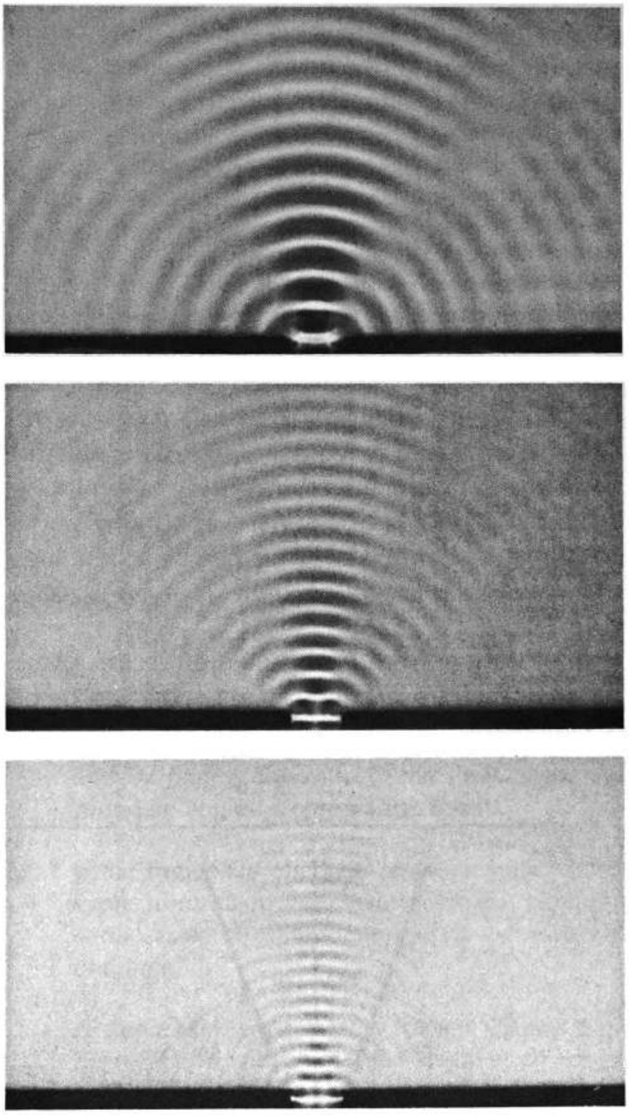
wavelength λ is shown in Fig. 7 – 19. In the middle of the pattern immediately beyond the opening, the wave crests are almost straight, but at the sides they curve, giving the impression of circular waves originating from the edges of the opening. This means that after passing through the opening not all of the wave propagates in its original direction. Part of it is bent. This phenomenon is called *diffraction*.

The diffraction of waves makes it hard to understand how we can explain the straight-line propagation of light with a wave model. If light is a wave, when it passes through a small hole some of the light should bend instead of traveling straight ahead. In a wave model of light there must be at least one great difference between light waves and water waves.

What could this difference be? We know the wavelength λ of the water waves. We do not know the wavelength of light waves; it may be very different. Let us, therefore, examine the diffraction of water waves of different wavelengths. In Fig. 7 – 20 we see three pictures of periodic waves with different wavelengths, each passing through the same opening. We notice a definite trend. In the picture at the top the wavelength is seven-tenths of the width *w* of the opening. There the part of the straight wave which gets through the opening is almost entirely converted into a circular wave. Or, in other words, the opening acts like a source of circular waves when straight waves fall on it. In the middle picture λ is five-tenths of *w*. In this case the wave which gets through is not so curved as in the top picture. It has an almost straight section in the middle and part of it still bends at the sides. In the bottom picture λ is three-tenths of *w* and here the bending fades away close to the forward direction, and we obtain an almost sharp shadow. Provided we keep the width of the opening *w* fixed, a further decrease in λ makes the shadow still sharper. We can also keep λ fixed and change *w*; then we find that the amount of bending does not depend on λ and *w* separately but only on the fraction λ/*w*. To sum up, waves are strongly diffracted when they pass through an opening of size comparable to their wavelength, and there is hardly any diffraction if the wavelength is very small compared to the width of the opening.



7–19 Straight waves passing through an opening. Note the curving of the waves around the ends of the barriers.



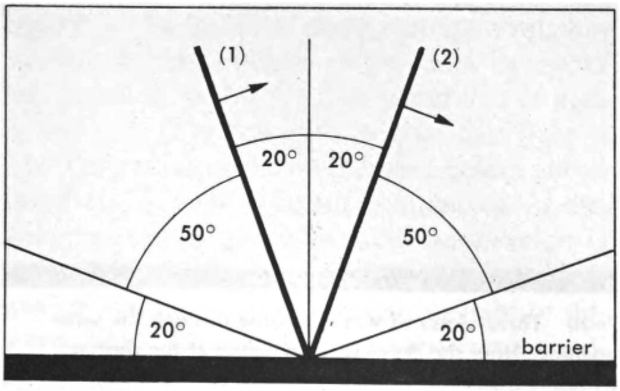
7–20 Three views of waves passing through the same opening. Note the decrease in bending at the shorter wavelengths.

We all know that light passing through a keyhole is not bent but seems to continue in its original direction. If light is a wave, this implies that its wavelength must be much less than the size of a keyhole. But to be certain that light is a periodic wave we should be able to do an experiment which will show diffraction.

Examine a light source through a narrow slit between two fingers. When the slit is half a centimeter across, a long, thin light source viewed through it appears normal. As the slit is narrowed to a width of about a tenth of a millimeter, however, the source appears spread out in a direction perpendicular to your fingers. Moreover, you can hold your fingers in any direction and the spreading out always appears perpendicular to them. This is a clear indication that the light is diffracted by the opening.

We can get a little more information from our experiments with light if we return for a minute to Fig. 7 – 20. Examination of Fig. 7 – 20 suggests an explanation of the fact that all shadows, even those of large obstacles, are slightly blurred by diffraction. Even in the bottom picture, where λ/w is about 3/10, the wave crests near the edge of the shadow are slightly bent. If we observe these waves at a distance far from the slit, the effect of this bending will be to spread the region reached by the waves. Since, as in the experiment with our fingers, we only begin to see the diffraction of light when the aperture is about one-tenth of a millimeter, we can now conclude that the wavelength of light is even smaller than this.

Briefly then, from our discussion in this chapter, light can be described as waves of very small wavelengths.



7–21 For Question 3.

QUESTIONS

1.* If a stick 10 cm long were placed horizontally in the ripple tank shown in Fig. 7 – 2, about how long would its shadow be?

2.* In Fig. 7 – 7, if $i = 25°$, what is the value of r'?

3.* In Fig. 7 – 21, the heavy black lines are crests; the arrows represent the direction of propagation of the pulses.

(a) Which is the incident pulse and which is the reflected pulse?

(b) What is the angle of incidence?

4. A straight pulse approaches a barrier at an angle of 30°. What is the direction of motion of the pulse after reflection? Indicate it on a diagram.

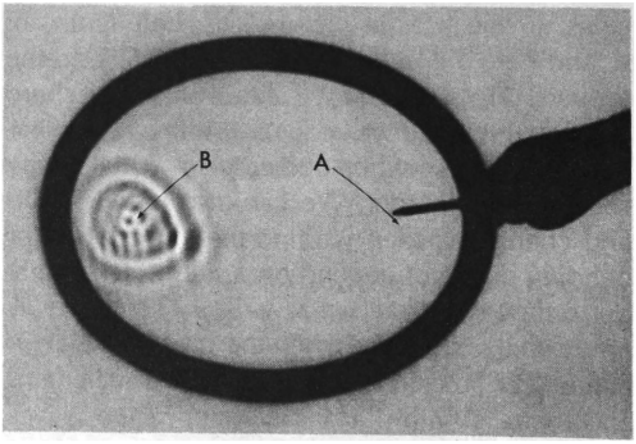
5. In Fig. 7 – 9, where is the object source? Where is the image source?

6. Describe the wave motion that results when you dip your finger into the center of a circular tank of water. What would be the motion under ideal conditions?

7. Suppose we place in a ripple tank a barrier in the shape of an ellipse as in Fig. 7 – 22. When a circular pulse is generated at point *A*, it reflects from the barrier and converges at point *B*.

(a) From this experiment what can you say about the geometry of an ellipse? (Hint: consider tiny segments of the circular pulse originating from *A* and see how the ellipse must be shaped so that all segments reach *B* at the same time.)

(b) What will happen if we generate a pulse at point *B*?

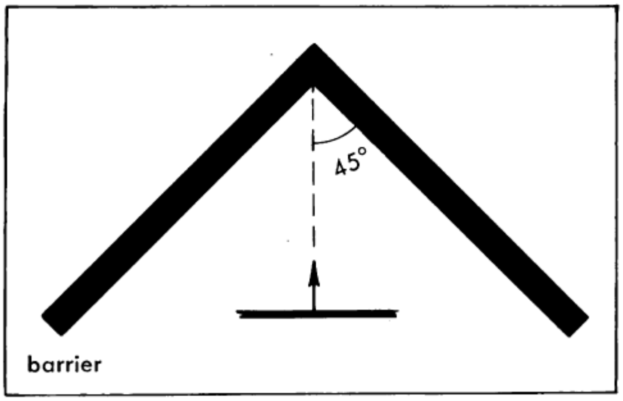


7–22 Reflection from an elliptical barrier. A pulse has been generated at A and is photographed as it converges on B.

(c) Will such a convergence also happen when you dip your finger in at some point other than *A* or *B*?

8. A circular pulse generated at the focus of a parabolic reflector is reflected as a straight pulse. What does this tell you about the geometric properties of a parabola? (Hint: consider tiny segments of the circular pulse originating at the focal point and see what must be the shape of the parabola to produce a straight pulse.)

9. In Fig. 7 – 23, a straight pulse approaches a right-angled barrier at an angle of 45°.
   (a) How does it reflect?
   (b) What happens if the wave is incident at some other angle?



7–23 A straight pulse approaching a right-angled barrier.

10.* A machine gun fires 10 rounds per second. The speed of the bullets is 850 m/sec.
   (a) What is the distance in the air between the flying bullets?
   (b) What happens to the distance between the bullets if you step up the rate of fire?

11.* If a wave generator produces 12 pulses in 3 seconds, what is (a) its period? (b) its frequency?

12.* Would you increase or decrease the frequency of the generator in a ripple tank in order to produce waves of greater wavelength?

13. (a) In a ripple tank, when one pulse is sent every $\frac{1}{10}$ sec, we find that λ is 3 cm. What is the speed of propagation?
   (b) In the same medium we send two pulses, the second one $\frac{1}{2}$ sec after the first. How far apart are they?

14. Assume you are looking at a periodic wave of frequency $f = 4$ per sec through a two-slit stroboscope. What do you expect to see if you rotate the stroboscope with a frequency of 1, 2, and 4 revolutions per sec?

15. A point source in the ripple tank produces circular periodic waves. By using a stroboscope to stop the motion, we measure the difference in radius between the first and sixth circular crests and find it to be 10 cm.
   (a) What is the wavelength?
   (b) Why did we not calculate the wavelength by using the radius, say, of the fifth pulse only?
   (c) Why do we use this method of measurement rather than take the difference between neighboring crests?

16. Circular waves are produced in the ripple tank by drops of water which fall at a constant rate. The wavelength is observed to be 1.2 cm. The experiment is repeated with the source moving at a uniform rate from one end of the tank to the other, along the center line. An observer near the end of the tank toward which the source is moving measures the wavelength and finds it to be 0.8 cm.
   (a) What is the ratio of the wave velocity to the source velocity?
   (b) What wavelength would an observer at the starting point measure after the source has moved some distance away from him?

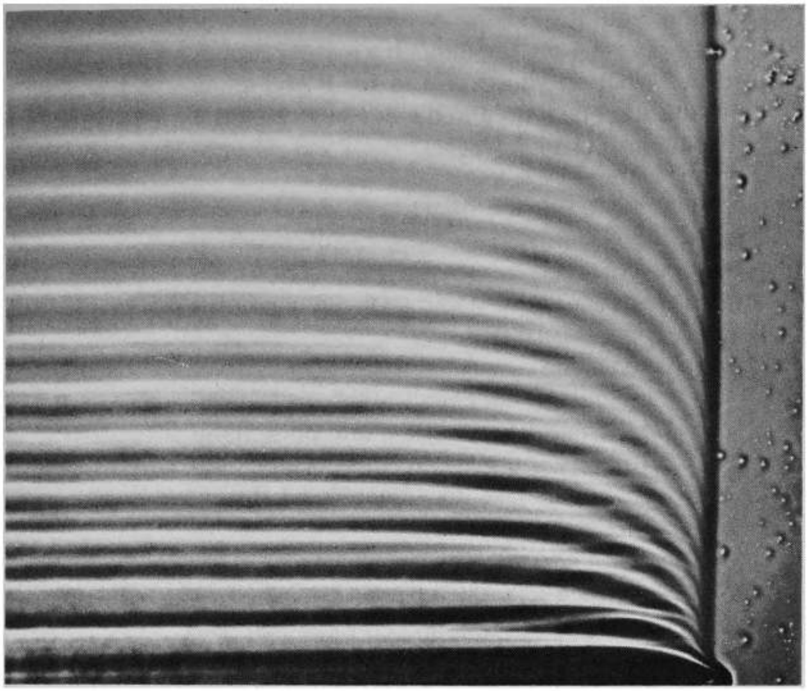
17.* In Fig. 7 – 13, where is the boundary between the area of deep and shallow water?

18.* What is the index of refraction in passing from the deep to the shallow water in Fig. 7 – 13?

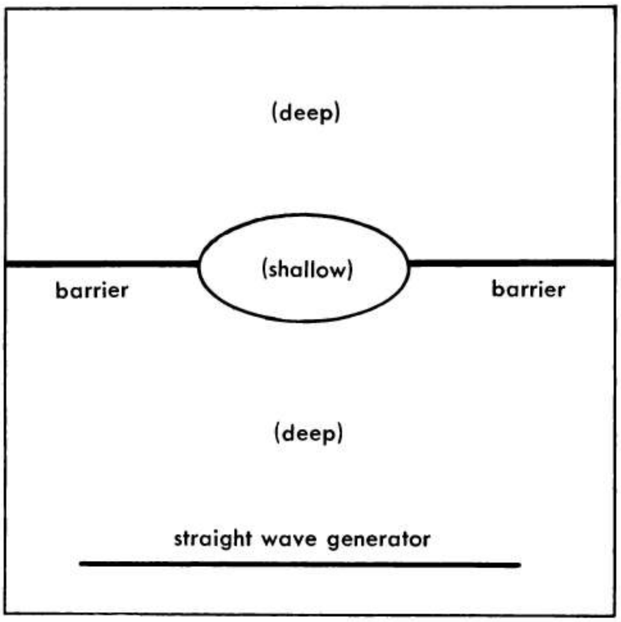
19. Measure the index of refraction in Fig. 7 – 14 by the method you used in the previous problem, and by finding the ratio of the sines of the appropriate angles. Compare the results.

20.* If the frequency changed when periodic waves went from deep to shallow water, would a stroboscope be able to stop all of the waves at one time?

21. A ripple-tank wave passes from a shallow to a deep section with an incident angle of 45° and a refracted angle of 60°.
   (a) What is the ratio of speeds in the two sections?
   (b) If the wave speed is 25 cm per second in the deep section, what is it in the shallow one?



7–24 **Curving of a straight wave when the water becomes more and more shallow from one side to the other.**



7–25 **The ripple-tank arrangement used in making the picture in Fig. 7–26. The oval region between the barriers has shallow water, while water in the rest of the tank is deep.**

22. (a) A tire on an automobile wheel has a circumference of 7.0 feet. When the wheel is turning 200 times per minute, what is the speed of the automobile in feet per min?

(b) A light wave whose frequency is 6.0×10^{14} per sec is passed through a liquid. Within the liquid the wavelength is measured and found to be 3.0×10^{-5} centimeters. What is the speed of light in this liquid?

(c) What is the wavelength in vacuum (from which the frequency was calculated)?

(d) What is the index of refraction of the liquid for light of this frequency?

23. The ripple tank is arranged so that the water gradually becomes shallow from one side to the other. Because of this, on one side of the tank the speed of a wave crest is different from that on the other side. As a result, straight waves become curved (Fig. 7 – 24). In the picture the pulses are moving toward the top of the page.

(a) Which is the shallow side?

(b) Does a similar phenomenon occur with light? Be prepared to discuss this in class.

24. Water waves traveling in the deep section of a ripple tank at 34 cm/sec meet a shallow part at an angle of 60°. In the shallow part all waves travel at 24 cm/sec. When the frequency is increased slightly, the waves are found to travel at 32 cm/sec in the deep section.

(a) Compute the angle of refraction for each case.

(b) Considering the ripple-tank conditions, is it easier to measure the two speeds and find their difference directly, or to measure it indirectly by the angular difference found in (a)?

(c) How can we detect small differences in the speed of light?

25. We set up the ripple tank as shown in Fig. 7 – 25 and generate a periodic straight wave. The resulting wave pattern is shown in the photograph in Fig. 7 – 26.

(a) Explain what is taking place.

(b) Of what optical arrangement is this a model?

26. In Fig. 7 – 27 we see a photograph taken under the same conditions as those in Problem 25 except that here the waves have a shorter wavelength.

(a) How do the two photographs differ?

(b) What can you conclude from this difference?

27. Suppose the oval region in Fig. 7 – 25 is deep and the surrounding region is shallow.

(a) What will happen to the straight waves?

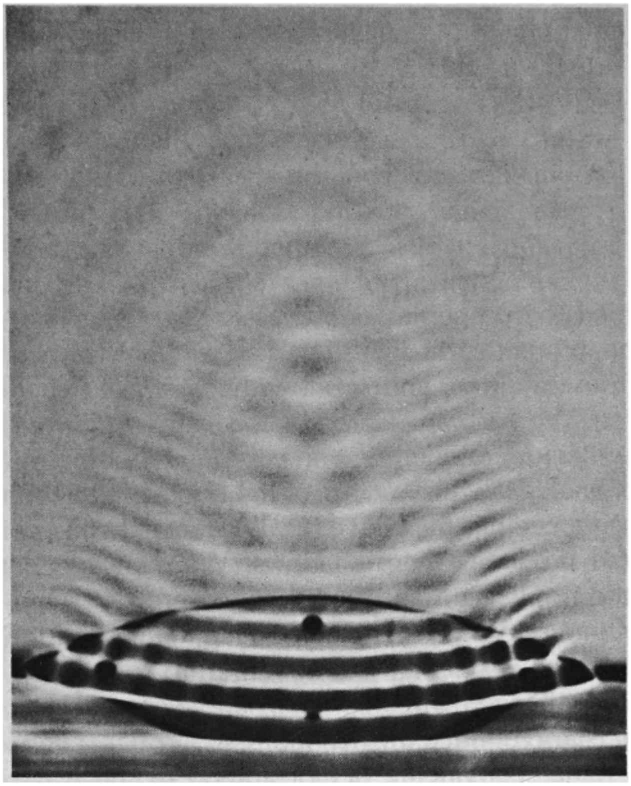
(b) Of what optical arrangement is this a model?

28. How do we know that the wavelength of light must be very much less than a centimeter?

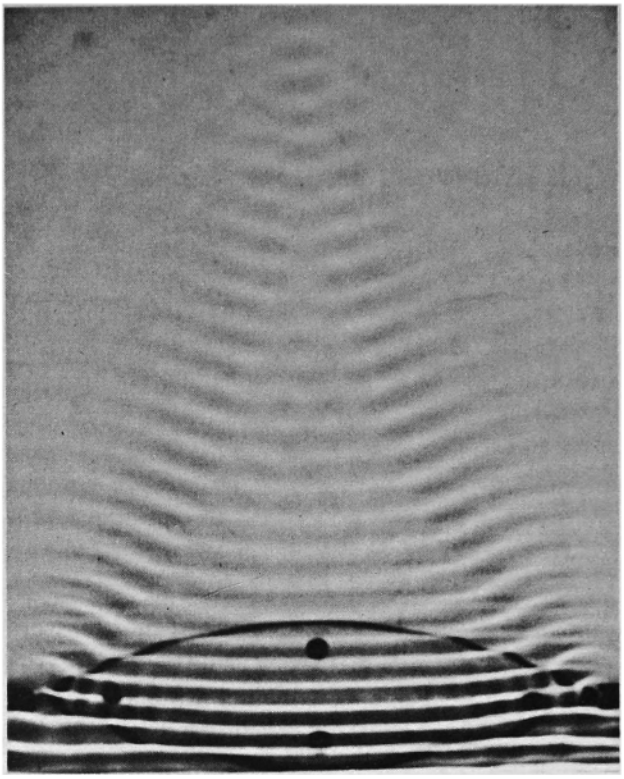
29. Imagine our eyes were sensitive only to light of wavelength 0.1 mm. How would this affect our ability to see? Could you thread a needle?

30. If sound is a wave phenomenon, how would you explain the common experience of hearing sound from around corners?

31. Sound waves in air usually travel at about 330 meters per second. Audible sounds have a frequency range of about 30 to 15,000 cycles per second. What is the range of wavelengths of these sound waves?



7–26 The wave pattern with periodic waves and an arrangement as in Fig. 7–25.



7–27 The same wave pattern as in Fig. 7–26 but with a shorter wavelength.

8

Interference

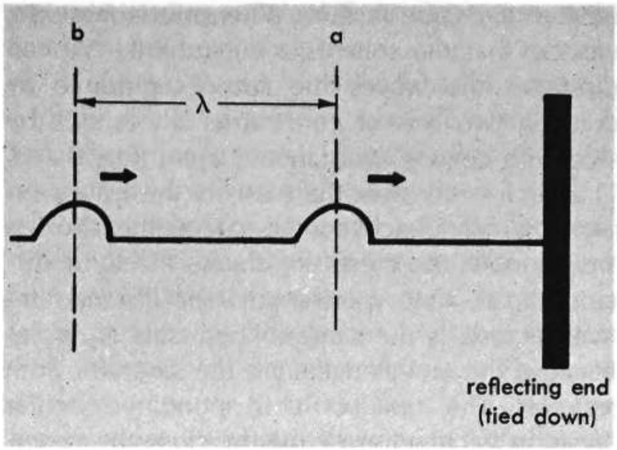
IN CHAPTER 7 we studied the properties of waves to see if they can account for the common properties of light. We had already learned how light is reflected from a mirror, and we found that waves obey the same law of reflection. We had also learned how light is refracted when it passes from one medium to another, and we found that the refraction of waves follows Snell's law just as that of light does.

In the last section of the preceding chapter we recalled that light propagates in straight lines and produces sharp shadows. We then experimented with waves in the ripple tank and found they were diffracted when they passed through an opening. We could reduce the amount of diffraction by decreasing the wavelength but could not completely eliminate it. Therefore we changed our approach; we went back to light, and found that light is also diffracted when it passes through a very narrow slit. This was a rather convincing demonstration of the wave nature of light. In this chapter we shall follow the latter approach. That is, we shall continue to study in detail the properties of waves and in the next chapter do the corresponding experiments with light.

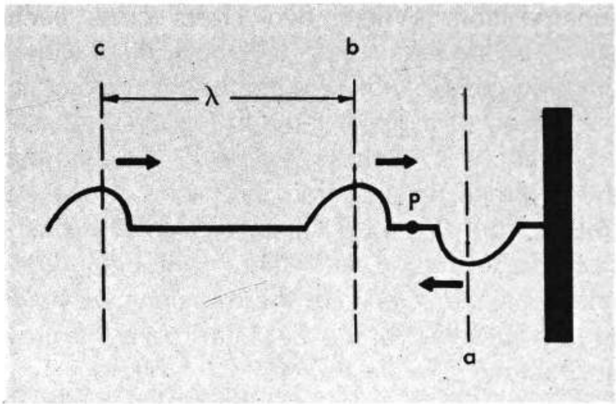
8–1 INTERFERENCE ON A SPRING

One of the most striking results of the experiments with waves on a spring described in Chapter 6 was that two pulses traveling in opposite directions passed right through each other. The shape of the displacement of the spring could be explained by adding the displacements of the individual pulses (the principle of superposition). For example, Fig. 6 – 9 shows the successive shapes of a spring as two opposite pulses pass through each other. The point *P* halfway between the two pulses remains undisturbed because adding the displacements of the individual pulses at this point produces a cancellation at every instant during the crossing. This behavior of two opposite pulses is somewhat difficult to see because you have to look at the pulses exactly at the moment they cross. However, if we use periodic waves, we can observe the cancellation more easily.

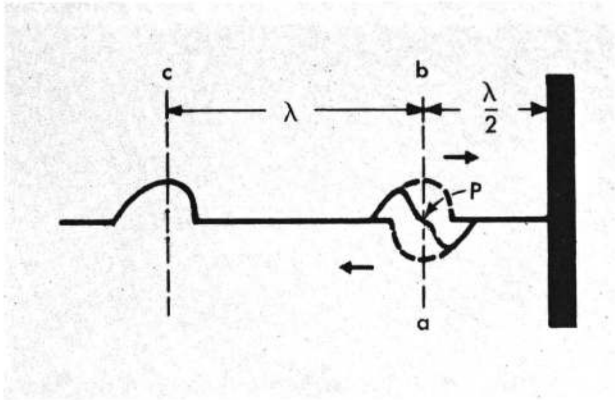
When a periodic wave travels along a spring which is tied down at one end, every individual pulse is reflected upside down. Now we know that a reflected pulse superposes with every oncoming pulse it meets. Thus, suppose we first consider only two of the pulses, *a* and *b*, a wavelength λ apart traveling toward the reflecting end (Fig. 8 – 1). Some time after the first pulse is reflected it will meet the second pulse and there will be a cancellation at the midpoint *P* between them (Fig. 8 – 2). In Fig. 8 – 3 the pulses are shown as they meet. Because they were originally a distance λ apart, we can see that the point *P* is a distance λ/2 from the reflecting end. The next pulse *c*, which reaches *P* later, will be superposed with the reflected pulse *b* so that the same cancellation occurs again. Because the wave is periodic, this will happen every time a pulse passes *P*, and, although the motion of the spring as a whole is complicated, the point *P* always remains at rest. We call such a point a *node*. There are other nodes spaced λ/2 apart, as you can see by working out where the next few must be (for example, the cancellation of *a* and *c*).



8–1 Two pulses travel toward the reflecting end of a spring.



8–2 The spring after one of the pulses has been reflected. The reflected pulse a is upside down and traveling toward the incident pulse b. A third pulse c is approaching at a distance λ behind b.



8–3 The pulses a and b of Figs. 8–1 and 8–2 meeting at point P.

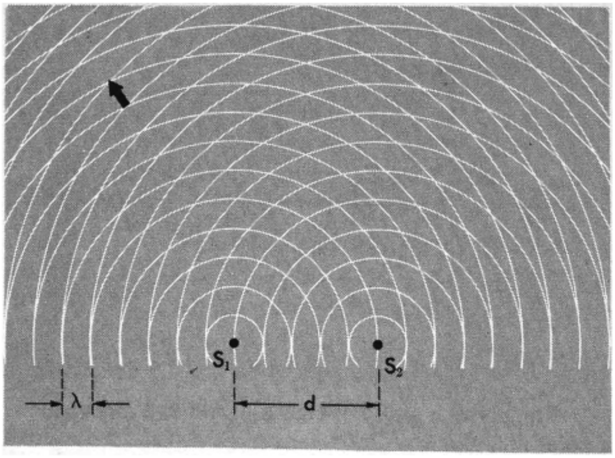
It is clear that we could have obtained nodes just as easily by sending appropriate periodic waves from opposite ends of a long spring. The use of the fixed end as a means of producing a wave moving in the opposite direction is only a convenience.

The phenomenon we have just described—that is, the superposition of two periodic waves to produce a series of nodes, is called *interference*. Rather than try to find the corresponding effect in light at once, we shall first go on to make a systematic study of interference between water waves in a ripple tank. Then we shall look for interference in light.

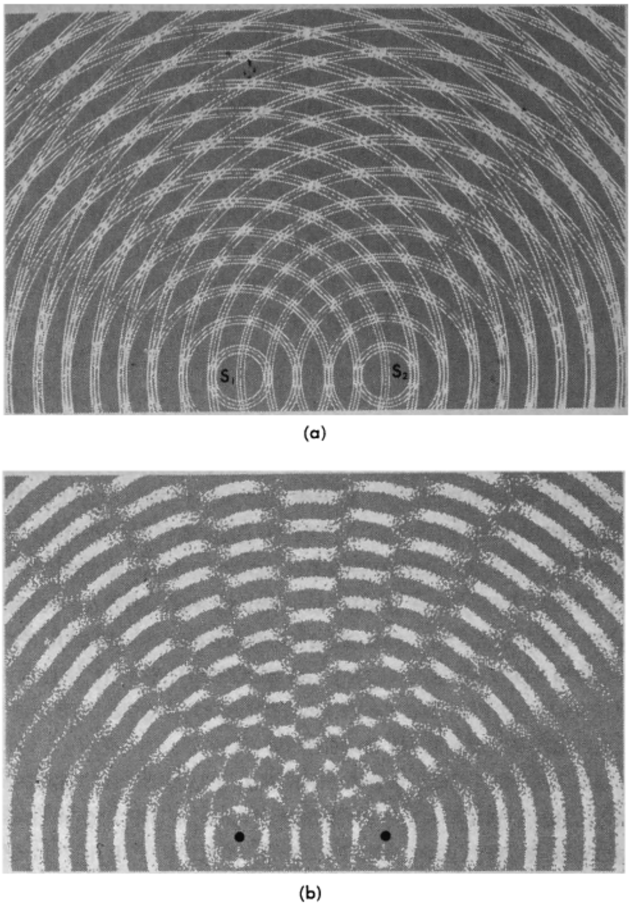
8–2 INTERFERENCE FROM TWO POINT SOURCES

For the purpose of studying interference in the ripple tank, we shall use point sources generating circular waves. Imagine two point sources side by side a distance *d* apart, each one generating pulses at the same frequency. Furthermore let them move so they dip into the water together—that is, so each source produces a crest at the same instant. When this is the case, we say that the sources are *in phase*. We can represent the waves the sources produce by drawing two sets of concentric circles side by side with centers a distance *d* apart (Fig. 8 – 4). The circles represent the crests of the waves expanding from each source. Since the sources are periodic, the crests are always the same distance apart—one wavelength. The distance between crests is the same in both sets of circles because the wavelengths are the same for both sources. The radii of corresponding circular crests in each set are equal because the generators are in phase.

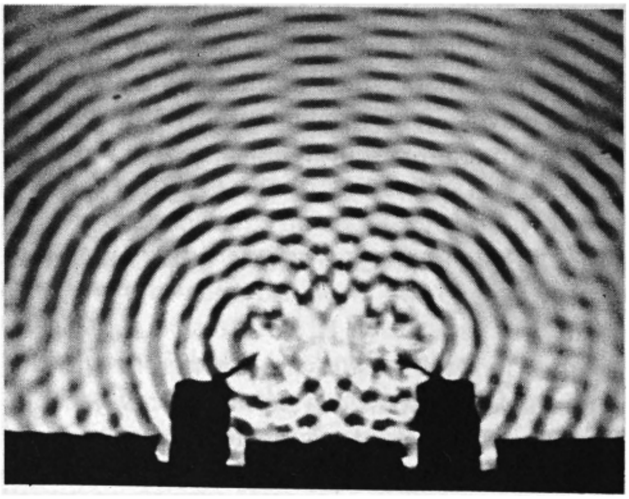
What will happen when the waves from the two sources overlap? Let us try to predict the resulting wave pattern by using the principle of superposition. Where two crests cross each other, a "double crest" will be formed. Such "double crests" will produce bright regions on the screen of a ripple tank. In Fig. 8 – 5 (a) we have emphasized these regions by brightening them. Where a crest from one source meets a trough from the other, the water will be practically undisturbed and there will be a gray image on the screen. In these regions we have used some dots in Fig. 8 – 5 (a) to give a gray appearance. Finally, where two troughs meet, a very dark image will be formed on the screen. In Fig. 8 – 5 (a) we have left these regions dark. In Fig. 8 – 5 (b) we have suppressed the construction lines, leaving only the pattern we expect to see. Therefore, the superposition of the waves in Fig. 8 – 4 should result in the pattern in Fig. 8 – 5 (b). An actual photograph of the waves from two point sources shows that this prediction is correct (Fig. 8 – 6).



8–4 The circles represent the crests of waves from two sources S_1 and S_2 a distance d apart. The sources are periodic and are in phase. The black arrow points out a region we shall examine more closely later (Fig. 8–7).



8–5 The pattern we predict by applying the principle of superposition to the waves from the two point sources shown in Fig. 8–4. Lighter areas show where crest meets crest; zigzag area is undisturbed water, where crest meets trough; places where troughs meet are left dark.

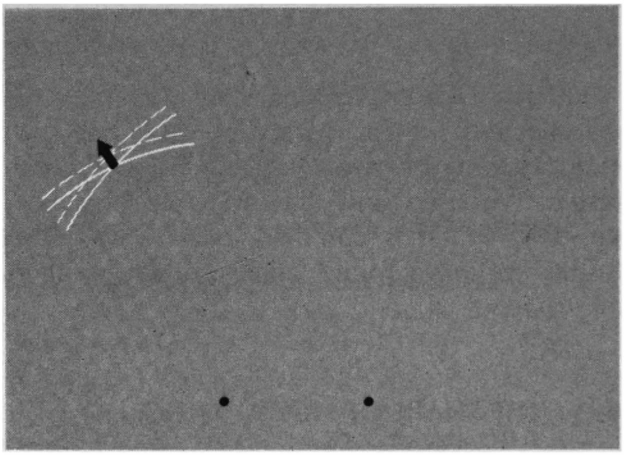


8–6 A photograph of the interference pattern from two point sources in phase. Notice the nodal lines radiating outward.

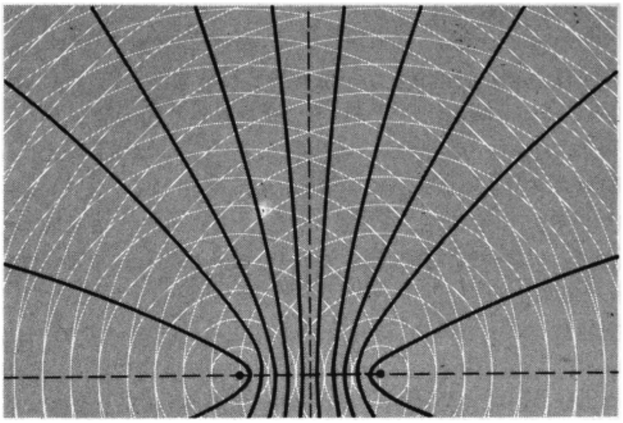
We have just constructed an interference pattern by superposing the waves from two sources. We have found that at one moment what we see agrees with the predicted pattern. Now let us consider how the waves in the pattern move. We shall start by finding out how one "double wave crest" moves. In Fig. 8 – 7 we show the two wave crests that cross to form the double crest at the tail of the arrow in Fig. 8 – 4. The dashed arcs in Fig. 8 – 7 represent the same two crests a short time later. Each crest has spread away from its source; and as a result the double crest moves away from the region of the sources in the direction of the arrow. If we wait for one whole period T, the wave crest from each source will have moved out a whole wavelength λ and the double crest will have moved from the tail to the head of the arrow in Fig. 8 – 4. The double crests and double troughs all over the ripple tank have the same kind of outward motion. Consequently, each row of double crests and double troughs moves away from the source region while new double crests and troughs are formed near the sources. Each row is a moving train of waves.

What happens in the gray regions, the regions between the moving wave trains? Here crests lie on troughs at all times and the water surface is not wavy. To see why, let us examine the line between two wave trains at any given instant. In Fig. 8 – 5 (a) pick one of the dotted gray regions. Start at the crest of a wave from S_1. Because it lies on a trough from S_2, there is no net displacement of the water surface. It is practically undisturbed. Now let us move out from the sources along this region. As we do, we descend from the crest of the wave from S_1 and we go up from the trough of the wave from S_2. As long as the waves are nearly symmetrical—so that a trough looks like an inverted crest—the addition of the displacements of the two waves continues to give zero total displacement. As we look farther out, we come to a trough from S_1 at the same place that we reach a crest from S_2. The upward and downward displacements still cancel; and we can see that by continuing out along the line through the intersections of crests and troughs, we shall find practically undisturbed water. Looking at Fig. 8 – 6 (or, even better, at the surface of a ripple tank), we can see these lines of undisturbed water extending outward from the region of the sources and separating the moving wave trains of reinforced crests and troughs. By analogy with the nodes on a spring, we call these lines of zero disturbance *nodal lines*. In Fig. 8 – 8 we have drawn them as thick black lines.

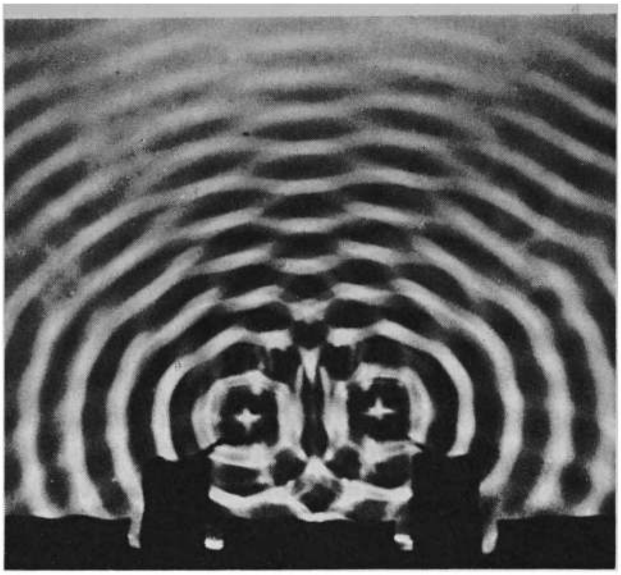
With different wavelength or different separation of the sources, the interference pattern changes in detail; but the general structure of such patterns is the same. Another pattern with a longer wavelength is shown in Fig. 8 – 9. Again you can see the nodal lines, and the waves moving out between them. The waves are slightly blurred because the film was exposed $\frac{1}{50}$ of a second, and in that time the waves moved an appreciable fraction of a wavelength.



8-7 The solid white lines show the double crests at the tail of the black arrow in Fig. 8-4. The dashed arcs show two crests a short time later. The two black dots represent the sources.



8-8 The nodal lines from two sources. Between the nodal lines are moving double crests and moving double troughs.



8–9 A two-source interference pattern like that in Fig. 8–6, but with a longer wavelength.

Although we considered water waves in our investigation, we really have not used any special property of water waves to obtain our result. We used only the principle of superposition, which is common to all waves. The results we have found apply equally well to all waves.

8–3 THE SHAPE OF NODAL LINES

If we examine Figs. 8 – 6 and 8 – 9 we notice that, although the nodal lines are slightly curved near the sources, they soon become quite straight. Another striking fact is that the number of nodal lines decreases as the wavelength increases.

In order to facilitate further discussion of these patterns, let us number the nodal lines so we will be able to refer to them conveniently. To do this we reexamine Fig. 8 – 8 and notice that the pattern is symmetrical—that is, it looks exactly the same to the left of the central (dashed) line as it does to the right. This should not be too surprising, as our sources also look exactly the same left and right. For this reason we need count only half the nodal lines, say those to the right. Thus we call the first one to the right of the central line the first nodal line. The next one is called the second nodal line, and so on. When we want to refer to one of the nodal lines and don't care which particular one it is, we usually say the *n*th nodal line, where *n* is an integer (1st, 2nd, 3rd, etc.).

Let us call a point on the first nodal line *P*, and connect it to the two sources by drawing the lines PS_1 and PS_2 (Fig. 8 – 10). We call these lines the path lengths from *P* to S_1 and from *P* to S_2. By counting the crests on the diagram it is easy to see that, for our particular point, $PS_1 = 3\lambda$ and $PS_2 = 2\frac{1}{2}\lambda$, so that the difference in the path lengths is

$$PS_1 - PS_2 = \tfrac{1}{2}\lambda$$

If we had taken any other point on the first nodal line, we would have found the same difference $\frac{1}{2}\lambda$ between the path lengths. Indeed, we may say that the first nodal line is composed of those points *P* for which the difference in the path lengths is $\frac{1}{2}\lambda$, so that one crest and one trough always arrive there at the same time.

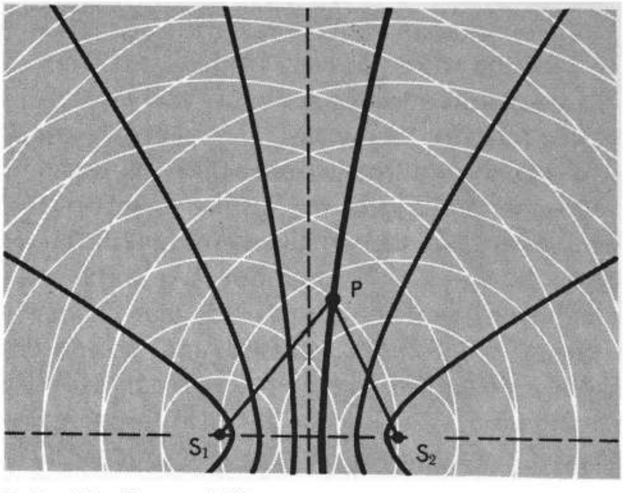
The second nodal line can be characterized in a similar way. In this case, if *P* is any point on the second nodal line, the difference in path lengths is

$$PS_1 - PS_2 = \tfrac{3}{2}\lambda$$

which can also be seen from Fig. 8 – 10. Continuing this procedure, we arrive at an equation describing the *n*th nodal line:

$$PS_1 - PS_2 = (n - \tfrac{1}{2})\lambda$$

According to this equation we can construct the nodal lines by finding the intersections of circles



8–10 The first nodal line. For any point P on the line the difference in path length from P to S_1 and P to S_2 is half a wavelength.

of radii r centered at S_2 with circles of radii $r + (n - \frac{1}{2})\lambda$ centered at S_1, where r has a different value at each point along the nodal line, while n is constant along any one line (Fig. 8 – 11).

8–4 WAVELENGTHS, SOURCE SEPARATION, AND ANGLES

In a ripple tank we can measure the path lengths to any point on a nodal line; and, using

$$PS_1 - PS_2 = (n - \tfrac{1}{2})\lambda$$

we can find the wavelength λ. We do not need to stop the waves to make such a measurement. The nodal lines stand still while we take our time measuring PS_1 and PS_2.

It is often convenient, or even necessary, to make our measurements at a point P which is far away from S_1 and S_2. But if we then measure the two large lengths PS_1 and PS_2 directly and subtract to find the small difference between them, we shall have a hard time getting sufficient accuracy. We may subtract away most of our measurement, leaving only our errors. We therefore look for a more accurate way to measure the path difference.

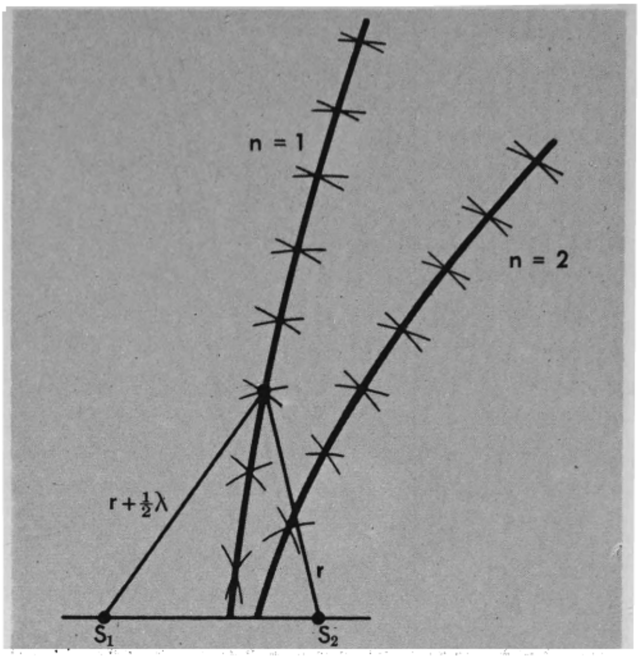
For any point P far away from the sources, the difference in the path lengths $PS_1 - PS_2$ depends on the angle between PS_1 and d. Consider Fig. 8 – 12(a), which shows the two sources S_1 and S_2, and a point P very far away compared to the source separation d. The distance PA is made the same as PS_2 so that the angles 1 and 2 are equal and $PS_1 - PS_2 = AS_1$. The farther away P is, the more nearly parallel the lines PS_1 and PS_2 become. We shall consider only points P that are far enough from S_1 and S_2 so that for all practical purposes PS_1 and PS_2 are parallel. Then we can draw Fig. 8 – 12 (b). Because angles 1 and 2 have become right angles, the triangle AS_1S_2 is a right triangle. Therefore

$$\frac{AS_1}{d} = \sin\theta$$

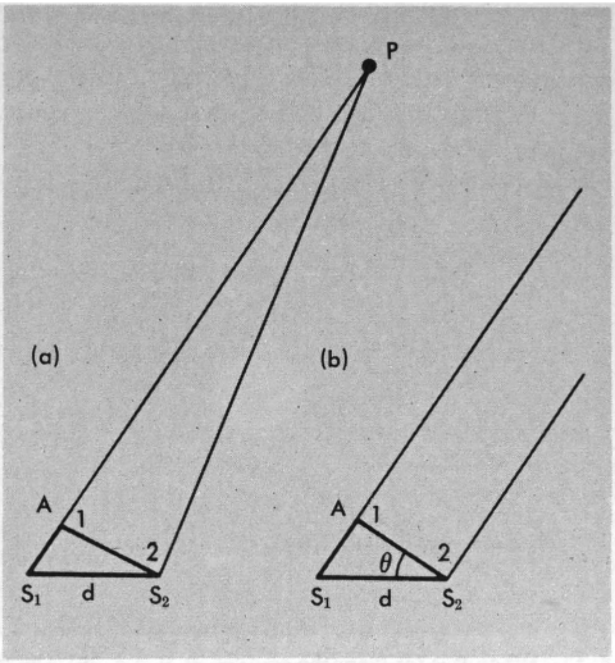
where θ is shown in the figure. Recalling that AS_1 is the path difference, we then find

$$PS_1 - PS_2 = d\sin\theta$$

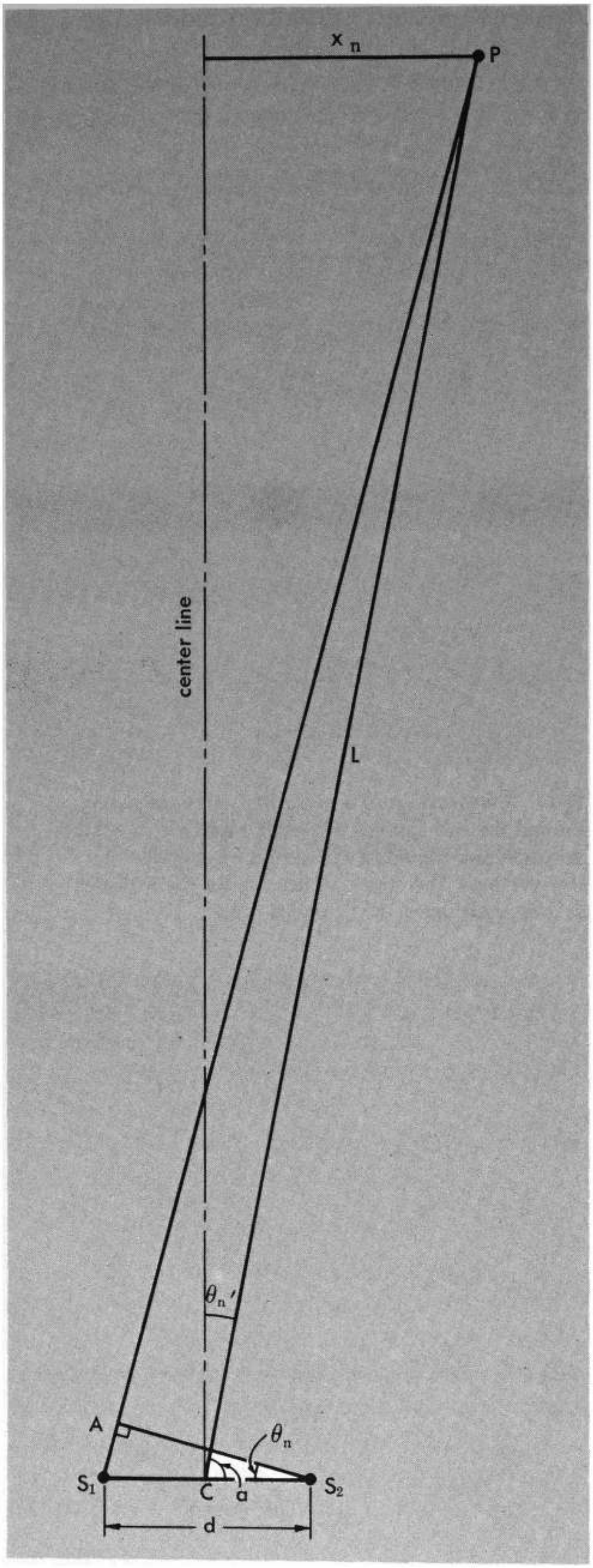
This equation expresses the path difference in terms of source separation and angle. The angle θ tells us the direction of P with respect to the sources. When $\theta = 90°$, for example, P is to the right on the line passing through the



8–11 Construction of a nodal line. Arcs are swung around the two sources with radii r and r + (n − 1/2)λ respectively. The nodal curve runs through the intersections. The curves shown are for the first and second nodal lines: n = 1 and n = 2.



8–12 (a) The path difference AS_1 can be determined in terms of the source separation d and an angle. (b) When P is far from the sources, $AS_1 = d \sin\theta$.



8–13 When P is far from the sources, $\theta_n + a = 90°$. Since $\theta_n' + a = 90°$, $\theta_n = \theta_n'$. Also, $\sin \theta_n' = x_n/L$, and so $\sin \theta_n = x_n/L$.

sources; when $\theta = 0$, P is in the direction of the top of Fig. 8 – 12(b).

Now when P is on the nth nodal line

$$PS_1 - PS_2 = (n - \tfrac{1}{2})\lambda$$

Consequently,

$$(n - \tfrac{1}{2})\lambda = d \sin \theta_n$$

or

$$\sin \theta_n = (n - \tfrac{1}{2})\lambda/d$$

as long as P is far away from S_1 and S_2. Incidentally, this result tells us that far from the sources the direction of a nodal line does not change. It is given by the angle θ_n. Far from the sources, therefore, the nodal lines must be straight as we noticed in the last section. Actually, if these straight portions of the nodal lines are extended back toward the sources, they pass through the midpoint of the line between the sources.

In the last section we also noticed that the number of nodal lines increases as the wavelength decreases. We can relate this observation to our equation for the direction of the nodal lines far from the sources. Since $\sin \theta_n$ cannot be greater than 1, then $(n - \frac{1}{2})\lambda/d$ cannot be greater than 1. The largest value of n which satisfies this condition is the number of nodal lines on each side of the center line. This number depends only on λ/d, and increases as λ decreases. We can make an approximate measurement of the wavelength, merely by counting the number of nodal lines.

To make an accurate determination of λ, we can find the direction of the nth nodal line, that is, the angle θ_n and calculate λ from the equation $\sin \theta_n = (n - \frac{1}{2})\lambda/d$. In the ripple tank, θ_n is easily found; this is not always the case with other waves. Therefore we shall look for a way to determine $\sin \theta_n$ directly without first measuring the angle θ_n itself. Let the point P in Fig. 8 – 13 be on the nth nodal line far away from the two sources S_1 and S_2, so that the lines CP and S_1P are practically parallel to each other and both are practically perpendicular to AS_2. Since the center line is perpendicular to d, we see than $\theta_n' = \theta_n$. But, from the figure, $\sin \theta_n' = x_n/L$ where L is the distance PC and x_n is the distance from P to the center line. Therefore we have

$$(n - \tfrac{1}{2})\lambda/d = \sin \theta = \sin \theta' = x/L$$

or $$\lambda = \frac{d(x/L)}{n - \frac{1}{2}}$$

where we have omitted the subscript $_n$ on θ, θ', and x, but you must remember that θ, θ', and x refer to the nth nodal line.

An example will show how simple the procedure is. Suppose we are working with sources 10 cm apart. We may pick a point P on the third nodal line, measure its distance L from the midpoint of d, and measure the distance x from the center line. Suppose we find that L is 80 cm long and x is 48 cm. Therefore, x/L s 0.6. We can now test the accuracy of our ratio by measuring values of L and the corresponding values of x for other points on the third nodal line. If all the points are far enough away, x/L remains 0.6.

Now because we are working with the third nodal line we must use $n - \frac{1}{2} = \frac{5}{2}$ and with $d = 10$ cm we find

$$\lambda = \frac{d(x/L)}{n - \frac{1}{2}} = \frac{10(0.6)}{\frac{5}{2}} = 2.4 \text{ cm}$$

Using the various nodal lines we can get several evaluations of λ. Agreement between the values obtained gives us a check on our reasoning and on our measurements.

8–5 PHASE

Two generators with the same period are *in phase* when they always dip into the water together, producing crests at the same instant. However, it is possible for two sources with the same period not to be in phase. For example, one of the sources may dip into the water slightly later than the other, after a time delay t. Since the natural time unit for a particular periodic motion is its period T, it is convenient to measure the time delay between the dips of S_1 and S_2 in fractions p of the period. Thus we use $p = t/T$ to measure the delay. For example, if each source dips every $\frac{1}{6}$ second and S_2 always dips $\frac{1}{18}$ second after S_1, then the fraction p is one third.

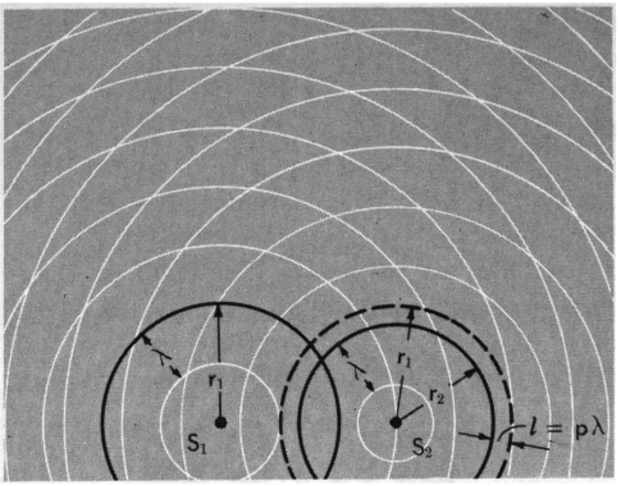
When two sources of the same frequency do not dip together we say that they are *out of phase*. The fraction p describes the *phase* delay of one source with respect to the other. There are no delays which are longer than the period T because we always measure the delay of the second source from the most recent dip of the first source, and its dips come a time interval T apart. Consequently there are no phase delays greater than 1. The value of p is always between 0 and 1.

Let us now use two point generators of waves operating so that S_2 has a phase delay of p with respect to S_1. What will the interference pattern look like? We can again try to discover this graphically by drawing two sets of concentric circles representing the wave crests from each source. As in Fig. 8 – 4, the crests in each set are always one wavelength λ apart; however, this time the sources are not in phase, and the radii r_1 and r_2 of corresponding crests from the two sources are not equal (Fig. 8 – 14). The radii of the delayed crests from S_2 are smaller than those of the corresponding crests from S_1 by a distance l equal to the fraction p of a wavelength λ:

$$l = p\lambda$$

For example, if S_2 is delayed one third of a period, the circles centered on S_2 are smaller than the corresponding ones around S_1 by one third of a wavelength.

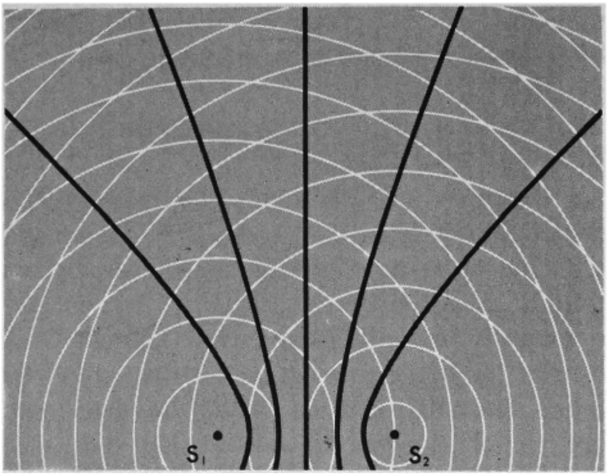
As an example let us see what happens when one of the sources is half a period behind the other. Then the distance l is half a wavelength, and the phase delay p is $\frac{1}{2}$. In Fig. 8 – 15 we have drawn the wave crests for such a situation and constructed the nodal lines by joining the points where a crest crosses a trough. We see



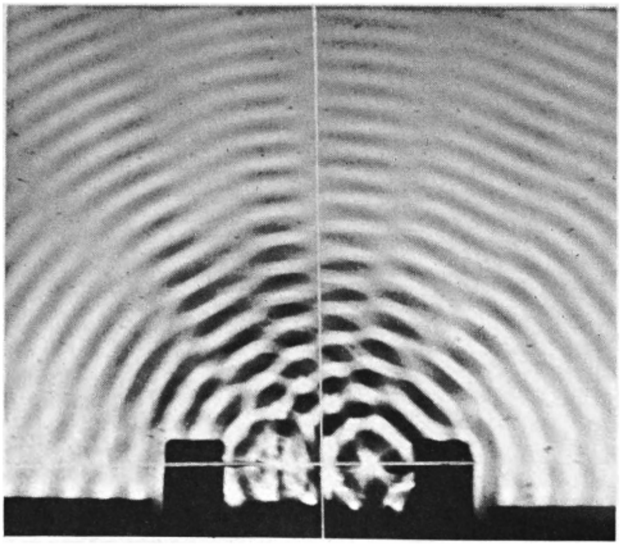
8–14 **The waves from two point sources out of phase. Source S_2 has a phase delay p with respect to S_1. The difference between the radii of corresponding crests is the distance $l = r_1 - r_2 = p\lambda$.**

that the pattern of nodal lines is different from the pattern for two sources in phase. For the same ratio λ/d the nodal lines are at the places where the reinforced crests used to be, and the reinforced crests are now where the nodal lines used to be. Compare Fig. 8 – 15 with Fig. 8 – 10. In each of these figures $\lambda = \frac{1}{3}d$, but in Fig. 8 – 10 the sources dip in phase while in Fig. 8 – 15 there is a phase delay $p = \frac{1}{2}$.

The photograph in Fig. 8 – 16 was taken with a phase delay $p = \frac{1}{2}$ between the sources. We can see, for instance, that there is now a nodal line along the center where in Fig. 8 – 6 (and Fig. 8 – 9) there are reinforced crests.



8–15 The pattern of nodal lines when there is a phase delay p = 1/2.



8–16 A photograph of the ripple-tank screen when two sources are operating with a phase delay p = 1/2.

We have given examples of the interference pattern for two particular choices of the phase delay, $p = 0$ and $p = \frac{1}{2}$. Actually we could have chosen any phase delay from 0 to 1, and in each case the interference pattern would have been different. The series of pictures in Fig. 8 – 17 show the interference patterns in the ripple tank for different phase delays of S_2. Changing the phase delay causes the whole pattern of nodal lines to shift in a definite direction. As the phase delay of S_2 increases, the radii of the crests from S_2 fall behind those from S_1 by an increasing distance $l = p\lambda$. Consequently the nodal lines bend more sharply around S_2, away from the center line. This is just what we see in the photographs. By including the distance $l = p\lambda$ in our computation of the difference in path lengths from any point on a nodal line to the sources S_1 and S_2, we can develop a general formula describing the nodal lines with any phase delay p. It is

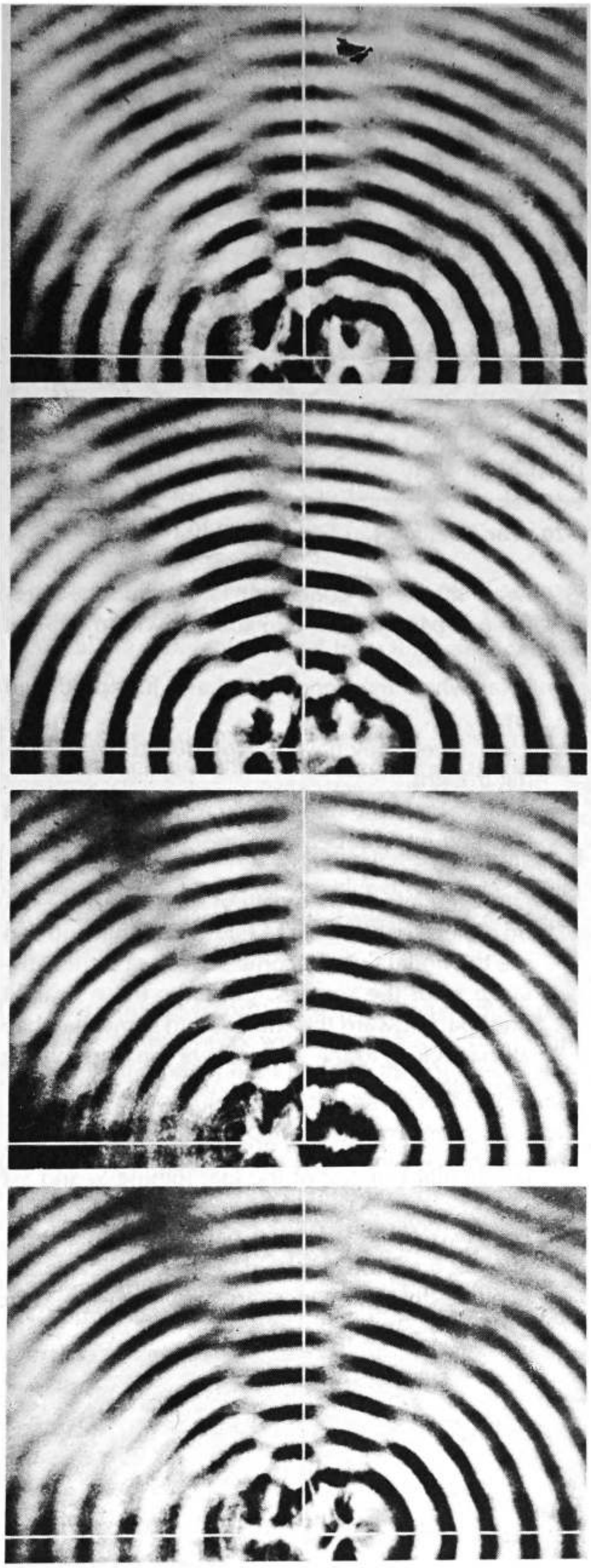
$$PS_1 - PS_2 = (p + n - \tfrac{1}{2})\lambda$$

and therefore (for the reasons given in Section 8 – 4) the angles at which the straight portions of the nodal lines are found are given by

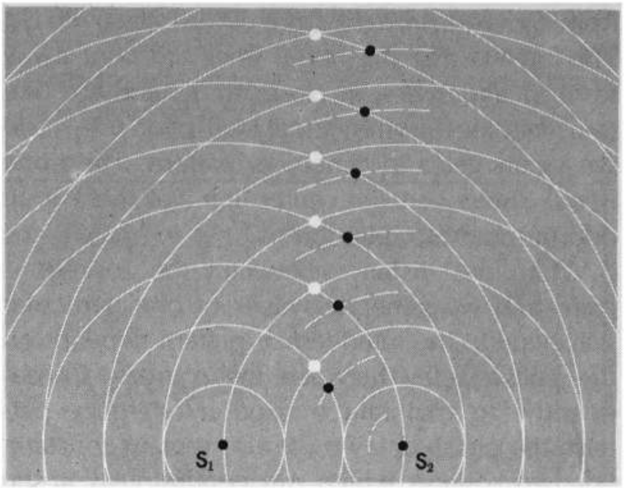
$$\sin\theta_n = (p + n - \tfrac{1}{2})\lambda/d$$

These equations can be understood by noticing that the nodal lines are determined by the intersections of arcs of radius $r + (n - \frac{1}{2})\lambda$ around S_1 with arcs of radius $r - p\lambda$ around S_2. At such points a crest from one source always arrives at the same time as a trough from the other. (See the end of Section 8 – 3.)

As the phase delay of S_2 changes, the wave trains of reinforced crests and troughs curl in the same way as the nodal lines. As the phase delay of S_2 is increased, any particular wave train bends away from S_1 and progressively curls around S_2. As an example, in Fig. 8 – 18 we have drawn the wave crests from two sources three wavelengths apart. The white circles from S_2 indicate waves generated when S_2 dips in phase with S_1. The dashed circles around S_2 indicate waves generated with a phase delay. Notice that the double crests bend around S_2 as the phase delay is introduced. In particular, with no phase delay there is a wave train along the center line perpendicular to the line through the sources. When S_2 is delayed, this wave train moves off to the right around S_2. All the other wave trains have a similar shift.



8–17 **The interference from two point sources with different phase delays p of the right-hand source. In the top picture the sources are in phase; p = 0. In the subsequent pictures p increases, passing through p = 1/2 in the third photograph.**



8–18 **The row of white dots represents a line of double crests formed when S_1 and S_2 are in phase. The black dots represent a line of double crests which results when S_2 has a phase delay. The delayed waves from S_2 are represented by dashed circles.**

8–6 SUMMARY AND CONCLUSION

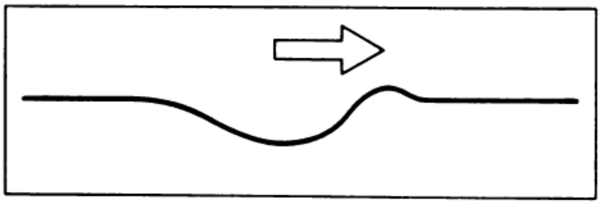
The interference pattern from two point sources in the ripple tank is characterized by a set of lines where the water surface remains undisturbed—the nodal lines. When the sources are in phase these nodal lines are distributed symmetrically about the center line, the perpendicular bisector of the line S_1S_2 joining the sources. In the immediate neighborhood of the sources the nodal lines are curved, but not very far away from the sources they become nearly straight. If the straight portions are extended back toward the sources, they pass through the center point between the sources. The number of nodal lines and the angle between any of these lines and S_1S_2 is determined by λ/d, the ratio of wavelength to source separation. This means that we can determine either d or λ if we know the other.

Even for the same wavelength and source spacing, different interference patterns are obtained for different phase delays p between the two sources. When the sources are in phase there is an even number of nodal lines symmetrically arranged about the center line. When the source S_2 produces crests later than S_1 with a time delay $t = pT$, crests from S_2 are all at smaller radii than the corresponding crests from S_1. The corresponding radii differ by the distance $l = p\lambda$. As a result the points where crests from S_2 cross troughs from S_1 are closer to S_2. This means that the nodal lines are bent away from S_1 and swing in closer to S_2; and the pattern becomes asymmetrical. Only for $p = \frac{1}{2}$ do we again get a symmetrical pattern. Then one nodal line covers the center line and there is an odd number of nodal lines.

When the distance between the sources is fixed and we generate waves of one particular wavelength, the interference pattern depends on the phase delay between the sources. Consequently, to maintain a fixed interference pattern, the phase delay must also remain constant. For two sources which run continuously at the same frequency, the phase delay will remain constant. But if each of the two sources is turned on and off in an irregular fashion, the phase delay will vary and with it the interference pattern. This shifting of the interference pattern will be of great importance in the next chapter, where we shall discuss the interference of light waves.

QUESTIONS

1.* If a periodic wave whose pulses each had the shape shown in Fig. 8 – 19 were traveling along a spring, and you wanted to produce nodes by sending appropriate periodic pulses in the opposite direction, what shape pulses would you use?

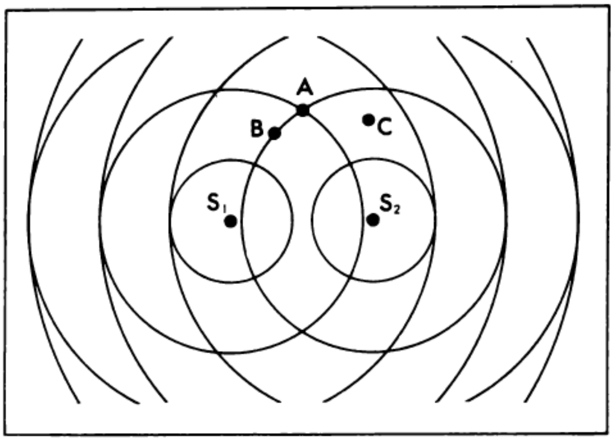


8–19 For Question 1.

2.* In Fig. 8 – 3, how far from the wall will pulse c meet pulse a to make the next node?

3. We showed in the text (Fig. 8 – 1, 8 – 2, and 8 – 3) that when pulses are incident periodically on the fixed end of a spring, the point P, a distance λ/2 from the end, never moves and is, therefore, a node. Extend the argument to show that the point P_2, a distance 3λ/2 in front of the end, is a node.

4.* In Fig. 8 – 20, the circles represent the crests of waves produced by sources S_1 and S_2. At which of the points A, B, and C is there a "double crest," a "double trough," or a nodal point?



8–20 For Question 4.

5. Draw the sets of concentric circles and the interference pattern from two sources with $d = 5\lambda$ at the time:

   (a) when the generators have just produced crests,

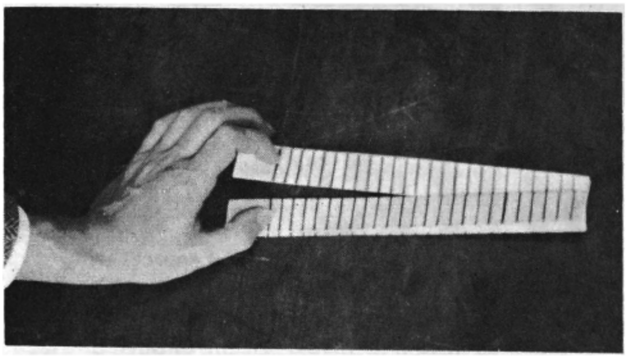
   (b) when they have just produced the following troughs.

   How have the reinforced crests moved during the time interval between these drawings?

6. Draw the interference pattern for the case $d = 5\lambda$ on a piece of paper large enough so that you can see the nodal lines become straight at a great distance from the sources. Continue these straight lines back toward the sources and show that they all pass close to the midpoint of the line joining the sources.

7.* In Fig. 8 – 4, is there a nodal line along the perpendicular bisector of the line joining S_1 and S_2?

8. Fold two pieces of ruled paper into long strips about 2 cm wide and hold them as in Fig. 8 – 21. Imagine that the lines are wave crests. Your



8–21 For Question 8.

fingers then represent the sources of the waves. Notice how the crests from both sources add together. Now by sliding the free ends sidewise, locate nodal lines and moving wave regions.

9. Consider an interference pattern produced by two point generators. What happens at the positions of the nodal lines if we place a third source exactly like the others at the point midway between them?

10.* Which nodal line in Fig. 8 - 10 is made up of all points P such that $PS_2 - PS_1 = 3\lambda/2$?

11.* You know the distances from a point on a nodal line to the two point sources in a ripple tank. What else do you have to know to calculate the wavelength of the waves?

12. Construct the nodal lines for two point sources with $\lambda/d = \frac{1}{3}$ by the method of Fig. 8 - 11. Is this really a different method from that used in Problem 6?

13. What would happen to the nodal lines if one of the two sources made weaker and weaker waves and then quit?

14.* For two point sources separated by a distance d is there a range of wavelengths λ which will produce an interference pattern with no nodal lines?

15. Two sources 6.0 cm apart operating in phase produce water waves with a wavelength of 1.5 cm. Draw the nodal lines far from the sources. Determine the position of each line by means of intersecting arcs of circles drawn from the two sources. Measure the angle between the second nodal line and the center line of the pattern. Compare the sine of this angle with $(n - \frac{1}{2})\lambda/d$.

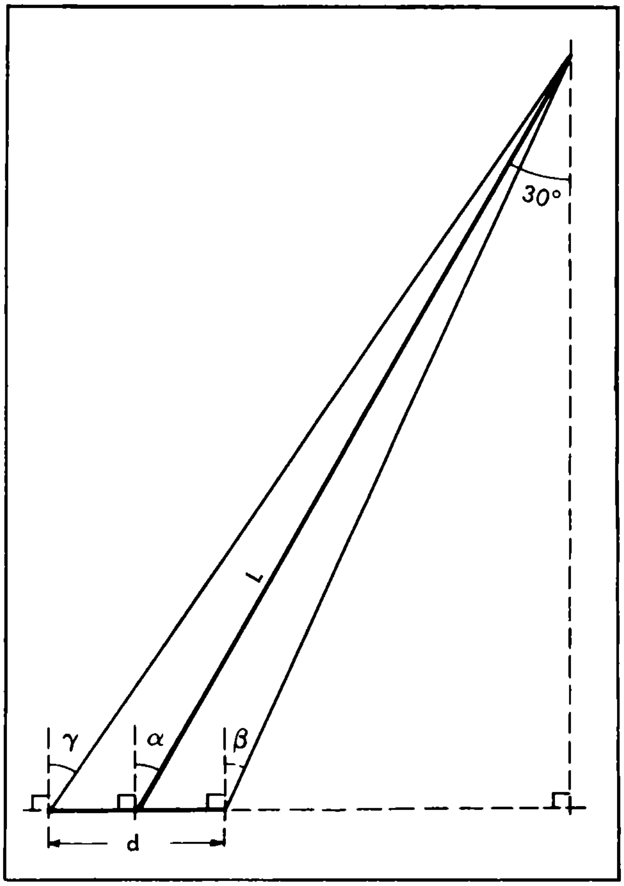
16. Suggest an interference experiment to prove that sound is a wave phenomenon. How could you use such an experiment to determine the wavelength of sound?

17. (a) From Fig. 8 - 6 and 8 - 9 find the ratio λ/d by using the equation $\sin\theta_n = (n - \frac{1}{2})\lambda/d$.

(b) The dimensions of the photographs are one fourth the actual size. By measuring d on the photographs estimate the actual value of λ.

18. Look up the definition of "hyperbola" and show that nodal lines are hyperbolas.

19. In Fig. 8 - 22, $L = 50$ cm, $d = 10$ cm, $\alpha = 30°$. What are γ and β? Find γ and β when $L = 500$ cm. Does this convince you that it is a good approximation to set $\gamma \approx \beta \approx \alpha$ when L is much greater than d?



8–22 For Question 19.

20. One red and one blue car are going around a circular racetrack 5.0 km in circumference. They move at constant speed. Each car takes 2.5 minutes for each lap. The blue car always comes around 0.50 minutes behind the red.

(a) What is the phase delay p of the blue car with respect to the red car?

(b) What is the speed of each car?

(c) If the track were only 4.0 km long, would this change the answers to (a) and (b)?

21.* In Fig. 8 - 14, $l = 0.25$ cm and $\lambda = 0.70$ cm. What is the phase delay?

22.* Is a phase delay of 0 the same as a phase delay of 1?

23. Show that for two sources with phase difference p, the first nodal line can be plotted from the equation: path length difference $= (p + \frac{1}{2})\lambda$.

24. Suppose we look at an interference pattern from a great distance L in front of the sources and find that the first nodal line is a distance x from the center line. If $x = .008L$ and $\lambda = .01d$, what is the phase delay of the sources? (See Fig. 8 – 13.) If $\lambda = .016d$, what is the phase delay?

25. Suppose that two point sources are generating waves with the same wavelength λ. They are placed in the ripple tank a distance $d = 5\lambda$ apart.

    (a) If the sources are in phase, what angle θ does the straight part of the first nodal line make with the central line?

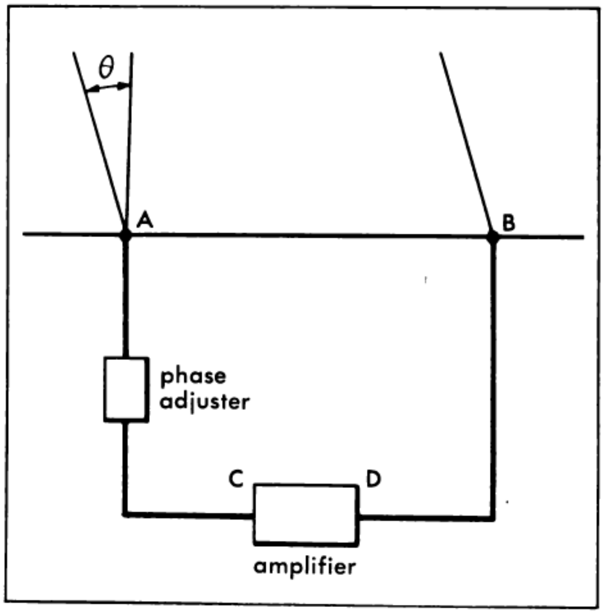
    (b) If the sources have a phase delay $p = \frac{1}{2}$ what is θ?

    (c) How many nodal lines will be produced?

26. In radioastronomy, one method for "looking" at different directions in the heavens is rather like using an adjustable-phase two point source rippler in reverse.

    Antennas A and B, tuned to the same wavelength, are set up a certain distance apart, and the signals they receive are mixed in an amplifier after the signal from one of them has been fed through a phase adjuster (Fig. 8 – 23). Maximum signal is obtained if the signals entering the receiver at C and D are in phase.

    The phase adjuster changes the phase by changing the length of the transmission cable between A and C. When the cable AC is equal in length to the cable BD, the signals from a point directly overhead will be in phase at the receiver. To "look" in the direction θ, the cable AC must be lengthened. If the distance AB is 100 m, and if the frequency to which the receiver is tuned is 20 megacycles/second, what phase difference must be introduced to make the system look in a direction 5° away from the vertical?



8–23 For Question 26.

27. Standard radio broadcasting stations operate on wavelengths of between 200 and 600 meters. The usual antennas are vertical towers that behave like point sources of waves in the plane of the earth's surface.

    (a) How could you arrange two point source radio antennas to give good coverage of a city built in a strip along the seacoast and waste a minimum of power over the water or in unpopulated areas?

    (b) If the city expanded inland, how would the station be able to provide better signal strength to the newly populated areas?

28. Two sources in a ripple tank are operating at frequencies of 15 cycles per second and 16 cycles per second. Describe the resulting pattern of nodal lines.

29. What will be the appearance of the interference pattern of two point sources if the phase delay between them is altered suddenly?

9

Light Waves

9–1 CAN WE SEE INTERFERENCE IN LIGHT?

In the last chapter we studied the interference patterns produced in a ripple tank by two point generators. Now we wish to do similar experiments with light, in order to see whether it has all the properties of periodic waves.

In designing an interference experiment with light we must keep in mind some important differences in the way we observe water waves and light waves. We can see water waves by standing at any place where light reflected from the waves reaches our eyes. We easily recognize a nodal line in the ripple tank because we can see where the water is undisturbed. Imagine now that we replace the two point generators by two light sources. How can we find out whether there are places where the light waves from the two sources cancel each other, places corresponding to the nodal lines in the ripple tank? In the ripple tank we could see the waves, but we cannot look across a light beam and see the light waves. (Recall Fig. 3 – 4.) To see this light, we shall have to get our eyes directly in the path of the light, or else place a reflecting screen in the path and observe the light that is reflected from it.

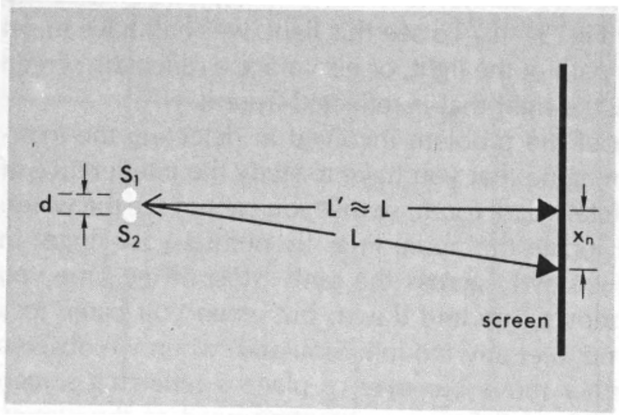
You can get an idea of the problem involved in detecting the interference of light if you imagine that you have to study the interference of water waves in a completely dark room, so that you cannot see the water. In this case you could locate the nodal lines by putting your finger in the water and moving it slowly across the tank. Most of the time you would feel the waves moving up and down; but when you came to a nodal line, you would not feel any motion. Similarly, when we observe light waves, we can either move our eyes or place a reflecting screen in the path of the light. Where light waves are reinforced on the screen, we can see light. Where a nodal line intersects the screen, we see a dark region.

At what angles do we expect to see these bright and dark bands? We learned from the ripple tank that, in the interference pattern generated by two point sources, the angles between the nodal lines depend on λ/d. For a given wavelength λ, these angles all increase as the source separation d decreases. We already know that the wavelength of light waves is much less than a tenth of a millimeter (Section 7 – 7); therefore, to achieve an observable separation of the nodal lines at a reasonable distance from the sources, we must place the sources close together. This means that they must be very small.

Also, to observe clear nodal lines we need sources which emit waves of a definite wavelength or at most a small range of wavelengths; otherwise, the nodal lines corresponding to one wavelength will be buried in the crests resulting from the others. Imagine, for instance, that each source emits waves of the two different wavelengths which give the patterns shown in Fig. 8 – 6 and 8 – 9. The total pattern you would then observe would be a combination of the patterns shown in the two photographs, and there would be no places which remain undisturbed in the total pattern.

In the following discussion we shall assume that the light we observe contains a narrow enough range of wavelengths so that the nodal lines, although they may become somewhat fuzzy, are not wiped out.

Imagine that we have two very small light sources placed a small distance d apart and that a screen is placed at a large distance L away from them, as shown in Fig. 9 – 1. This arrangement seems similar to that of a ripple tank in which there are two neighboring sources of circular waves. Consequently, from the results described in Sections 8 – 4 and 8 – 5, we might expect to see a pattern of bright and dark bars on the screen. The bright bars should be the regions where light waves from the two sources reinforce, and the dark bars should occur where the waves cancel. Just as in the ripple tank, if the two sources emit waves in phase there should be a bright region (that is, a large wave) hitting the middle of the screen. On each side of this central bright region, there should be dark bars where the first nodal lines meet the screen; and as we go away from the center, bright and dark bars should alternate.



9–1 This apparatus resembles the setup for producing interference in a ripple tank. But with two separate, ordinary sources of light S_1 and S_2, we cannot find interference. No pattern appears on the screen. Using the special, phase-locked sources of light called lasers, this arrangement can be made to show a pattern (Section 9–3).

In this pattern of alternating bright and dark bars, the nth dark bar on each side should be a distance $x_n = (n - \frac{1}{2})L\left(\frac{\lambda}{d}\right)$ from the center of the pattern, if x_n is much smaller than L, so that $L' \approx L$. Consequently, the spacing between neighboring dark bars should be $\Delta x = L\left(\frac{\lambda}{d}\right)$. Also, as in the ripple tank, if one source emits its waves with a phase delay p compared with the other, the pattern should shift off center. It should shift sidewise on the screen by the fraction p of the distance between the dark bars.

Can we observe these interference patterns? We already know that if light is a wave, its wavelength must be much smaller than a tenth of a millimeter. Consequently, $\left(\frac{\lambda}{d}\right)$ for any practical sources will be a very small fraction. However, we can choose a very large distance L to the screen; therefore, it should be possible to see the bright and dark bars.

In reality, when we try the experiment we have just described we do not see an interference pattern. No matter how we vary d or L, the screen is always uniformly illuminated.

Does our failure to observe an interference pattern prove that the wave model of light is a failure? Not necessarily. On the basis of the evidence we have discussed so far, the wave model may have failed, but another explanation is also possible. As we have just mentioned, the interference pattern produced by two sources depends on λ, on d, and also on the phase delay p. If the phase delay between the sources changes rapidly, the nodal lines and the dark bars on the screen must shift position rapidly.

Our eyes cannot follow these rapid shifts, so the screen will appear uniformly bright. In other words, the interference patterns may possibly be produced, but we fail to see them because they move too rapidly.

9–2 INTERFERENCE OF LIGHT WAVES: YOUNG'S EXPERIMENT

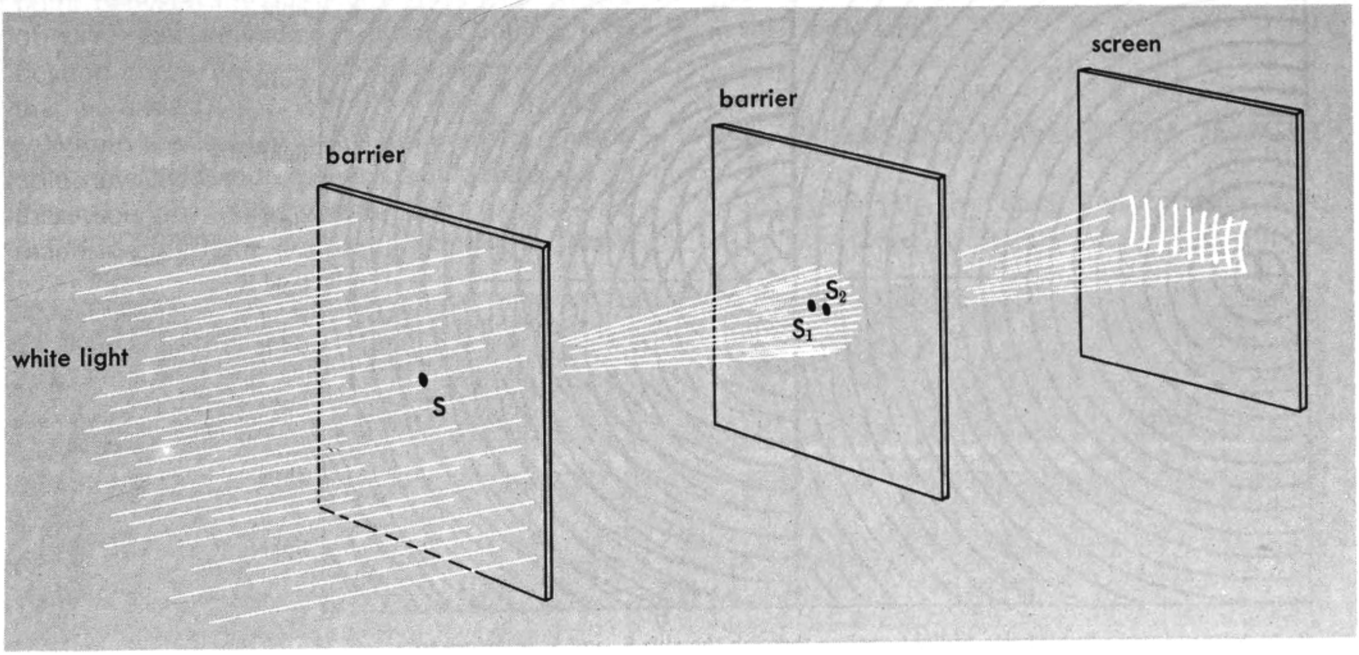
We can show that our inability to observe interference in the experiment just described is due to the rapid shifting of phases. To do this, we resort to a trick introduced in 1801 by Thomas Young. He found a simple way to lock together the phases of two light sources so that any interference pattern that is formed will not rapidly move away. The trick is to start with a single light source and split the light from it into two parts in phase with each other. These two parts, which act as if they came from two sources locked in phase, are then allowed to come back together to interfere with each other. With Young's method, we do indeed find the expected interference pattern on a distant screen.

Young's experimental arrangement is shown in Fig. 9 – 2. He used sunlight passing through a pinhole as a single source of light. The light spreading out from this pinhole source fell upon an opaque barrier which contained two pinholes, placed very close together and located equidistant from the source. Light originating from the pinhole source at any moment passed through the other two pinholes at the same time; the light at these two pinholes was then always in phase. With such an arrangement, the interference pattern of the light emerging from the pair of pinholes did not shift; it could be observed.

A permanent record of the interference pattern produced by this system can be made by putting a camera in place of the white screen to record the result on film. Such a picture is reproduced in Fig. 9 – A opposite page 132.

9–3 THE PHASE OF LIGHT SOURCES: ATOMS

We can couple two generators in a ripple tank with any desired phase delay, start them, and let them run as long as we wish. We do not have such control over light sources. That is shown by our inability to observe interference between light waves unless Young's method (or some similar method) is used to lock the phase of the two sources. We can also understand why the phase delay between two sources usually shifts rapidly. Consider two separate

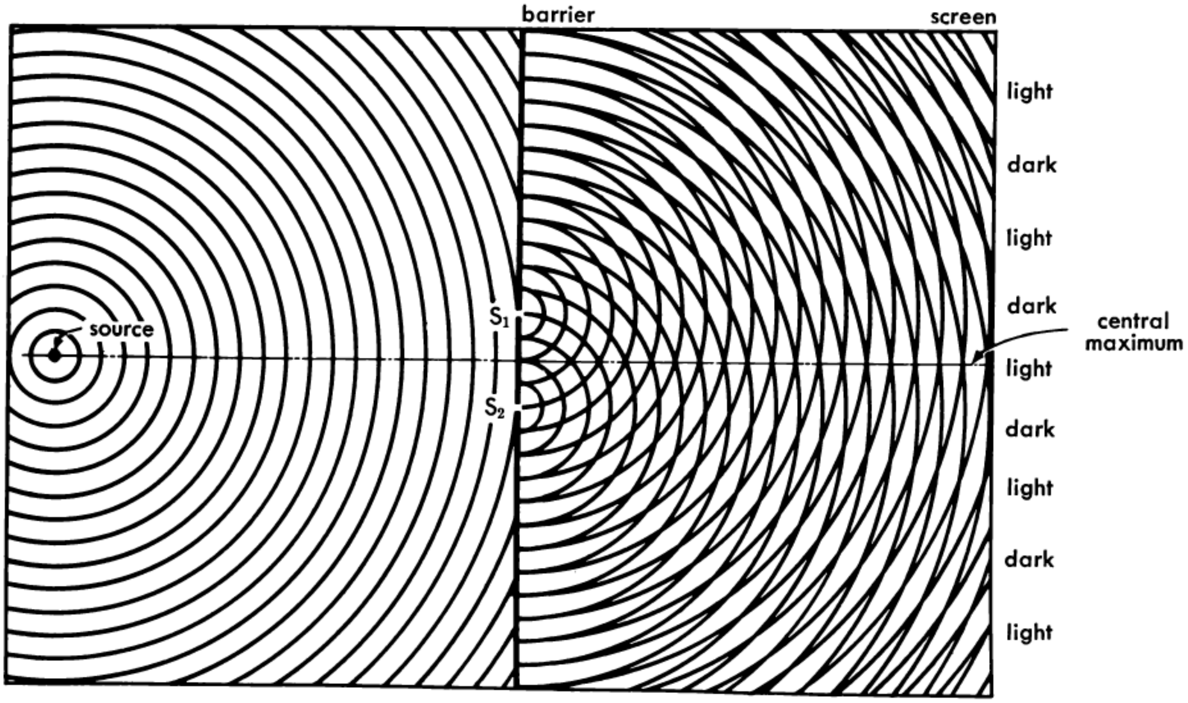


9–2 Young's experiment. The sunlight reaching the screen from the pinholes S_1 and S_2 all comes through the pinhold S. An interference pattern is visible on the screen.

sources of light. The light from each source comes from a large number of individual atoms; each atom sends out a burst of light waves only during a very short time. When we turn on the sources we start the overall process of the emission of light, but we do not control the individual atoms; they emit their light waves at random.

Today we can do this experiment more easily. We replace the sunlight and pinhole source by an incandescent light bulb having a long, straight filament; and we replace the two pinholes in the opaque barrier by the long, narrow slits. The slits must be very close together (about $\frac{1}{10}$ mm apart) and placed with their long dimension parallel to the line filament. Figure 9 – 3 is an idealized diagram of the wave pattern producing the interference, looking at the line source and slits from above. The similarity to a ripple-tank pattern is very clear. If you hold the page at the level of your eye and sight from the right edge of the figure toward the source, the nodal lines can be seen distinctly. These are regions of no wave disturbance and hence of no light; they have been marked "dark" where they intersect the screen. On the screen, we would see an interference pattern consisting of alternate bright and dark bars of various colors. The bars are parallel to the slits, with a bright one at the center, just as predicted by the wave model.

To see the significance of this situation for our interference experiments, let us consider a pair of atoms, one in each source. In all probability these atoms, considered as light sources, will be out of phase by some fraction p_1. The light waves from these two atoms will produce an interference pattern that depends on the value of p_1. A short time later these particular atoms will have stopped radiating, so we shift our attention to another pair of atoms which happen to be radiating at that moment. Since we have no way of influencing their behavior, these atoms will also be out of phase, but this time probably by a different fraction p_2. They will, therefore, produce a different interference pattern. Still later we consider a third pair of atoms, and so on. A typical time during which an atom emits light has been found to be about 10^{-9} sec. If we consider only one atom radiating at any moment in each source, we must change atoms every 10^{-9} sec to have the radiation continue. Then the interference pattern would also change in an irregular fashion every



9–3 Waves from a line source of light passing through the slits S_1 and S_2 interfere to give alternate bright and dark bands on a screen. In this idealized diagram, the slit width is less than the wavelength.

10^{-9} sec. Since this is certainly much too fast for our eyes to follow, we would see no interference pattern at all.

In reality there are many atoms radiating at the same time. At each moment the interference pattern will be determined by the superposition of light waves from all the atoms. Now let us concentrate on all the atoms that are radiating in one source at a given moment. They will have finished radiating in about 10^{-9} sec. By that time a new set of atoms will be emitting the light from that source. Consequently, the phase of the source will have changed at random just as if we switched from one single atom to another. This shifting of phase therefore goes on in each source in about 10^{-9} sec, and the interference pattern moves violently in this length of time. No wonder we cannot see any interference pattern.

When we use Young's method to obtain an interference pattern, the light wave emitted by each atom in a single source passes through two slits. These slits are so far away from the light source that the waves from all the atoms in the source arrive at the two slits with the same path difference (Fig. 9 – 4). As a result, all the light waves passing through the slits and going beyond them leave the slits with the same phase delay. In fact, if the plane of the slits is perpendicular to the line connecting the source and the midpoint between the slits, the slits act as sources of waves that are locked in phase. On the screen beyond them we see the interference pattern that we expect.

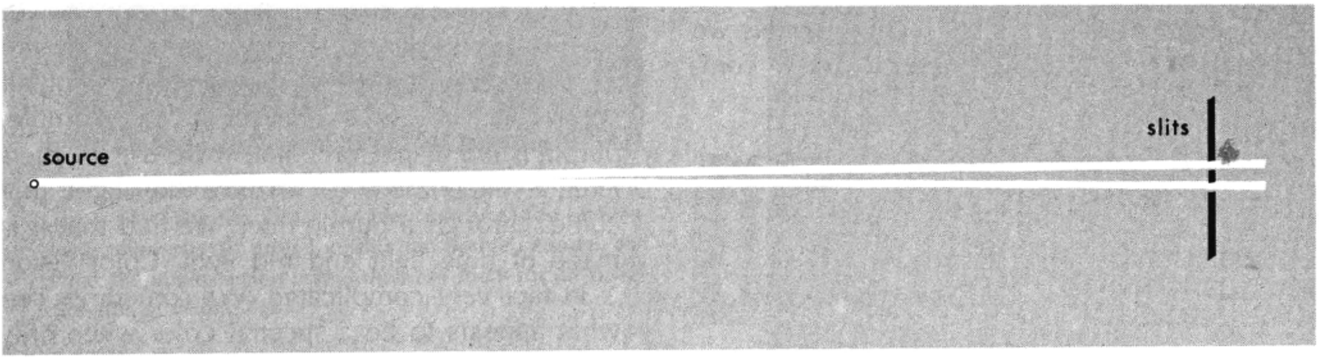
Within the last few years, a family of remarkable new light sources has been developed. In these sources, atoms within solids or gases are made to emit light, not at random but in almost locked phase, one atom tending to control another. The fixed phase condition can be maintained for periods thousands of times longer than the 10^{-9} second during which an individual atom in a hot filament emits light. These new devices are called lasers (for Light Amplification by Stimulated Emission of Radiation) and use a principle that can be applied in many forms.

In an experiment done in 1963, two independent lasers were made to work so well as phase-locked sources that light interference fringes were seen, in exact analogy to the interference pattern we see produced in a ripple tank by two independent wave generators. A technical difficulty prevented this from being an easy experiment, however. Although the phase-lock was good, the wavelengths could not be made precisely equal, which caused the position of the central fringe to drift rapidly across the screen, just as the ripple-tank pattern drifts when the two generators run at slightly different rates. The laser fringes were so fleeting that it was necessary to use high-speed television techniques to make a "snapshot" of the fringes at one position.

Physicists hope that one day it will be possible to make laser sources of identical wavelengths, fully locked in phase. Then interference effects with light will be as easy to see and understand as those with water now are in the ripple tank.

9–4 COLOR AND WAVELENGTH OF LIGHT

The photograph in Fig. 9 – A (opposite page 132) was taken by letting white light from a



9–4 Because the source is small and far from the slits, the path difference to the slits is very nearly the same for every atom in the source.

small source pass through two closely spaced parallel slits; the light spreading out from the two slits was brought back together again by the lens of a camera to form the interference pattern on the photographic film. Notice that the bright bar in the center (called the central maximum) is white, but the edges of the bright bars on the sides are colored.

We know that white light is a mixture of all colors, and the colored edges of the bars suggest that different colors are deflected at different angles in the interference pattern. The idea that different colors interfere at different angles implies that the wavelength of light is associated with its color. We previously had some reason to suspect an association of color and wavelength when we studied dispersion. Now we can settle the question by further interference experiments. Let us, therefore, look at the interference pattern of light of one color. In Fig. 9 – 5 we see the interference pattern for red light alone. It is a series of alternating bright and dark bars. The spacing of the bars in the photograph depends on the separation d of the two slits, the distance L from the slits to the photographic plate, and the wavelength of the light. Consequently, when we know the distances d and L, the separation of the bars in the photograph tells us the wavelength of the light. As we saw in Section 9 – 1, the separation Δx of neighboring dark bars is given by

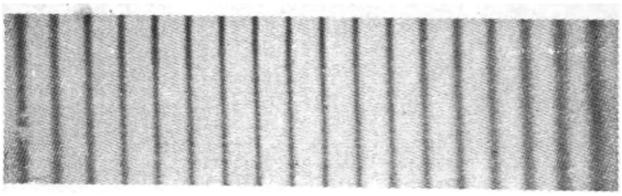
$$\Delta x = L\left(\frac{\lambda}{d}\right)$$

On rearranging this equation we get the wavelength as

$$\lambda = \left(\frac{d}{L}\right)\Delta x$$

By measuring across a large number of bars we can get an accurate value of Δx to use in computing the wavelength λ. For red light this turns out to be

$$\lambda_{\text{red}} = 6.5 \times 10^{-7}\ \text{m}$$



9–5 Interference pattern produced with red light. This is shown in color in the upper part of Fig. 9–B, opposite page 132.

We can go through the same procedure with light of any other color. If, for example, we use a blue-violet filter to select the light, we obtain a wavelength

$$\lambda_{\text{b-v}} = 4.5 \times 10^{-7}\ \text{m}$$

Other typical wavelengths, similarly measured, are given in Table 1. Our early conjecture that color is associated with wavelength is borne out by experiment.

TABLE 1

Wavelengths of Light in Vacuum

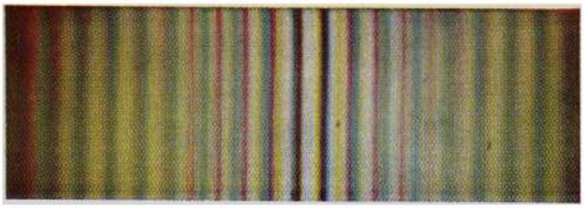
A. The wavelengths in angstrom units of the colors of the spectrum. The visible spectrum ranges from about 4000 Å in the deep violet to about 7500 Å in the deep red. (1 Å = 10^{-10} m)

Violet	<4500 Å
Blue	4500-5000
Green	5000-5700
Yellow	5700-5900
Orange	5900-6100
Red	>6100

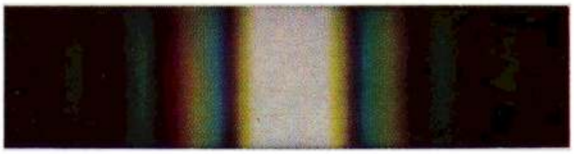
B. The wavelengths of some of the strongest visible lines in the spectra of a few common gases.

Lithium	Neon	Sodium
6103.6	5400.6	5890
6707.8	5832.5	5896
	5852.5	
Mercury	6402.2	Potassium
4358.4		4044
5460.7		4047

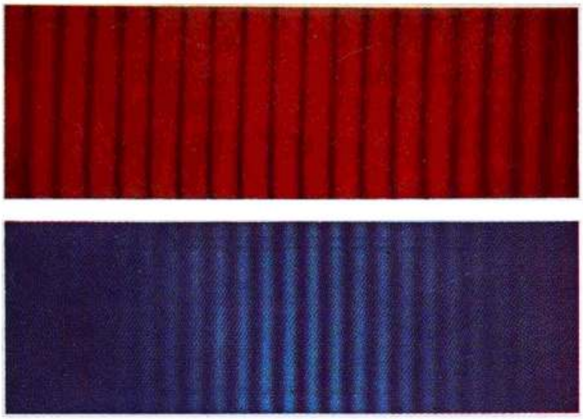
The colors associated with definite wavelengths are known as spectral colors. They are the same ones seen in the rainbow. Not all colors we see are of this kind. For example, purple is not a spectral color. If we use an interference experiment to analyze the light that comes through a purple filter, we find that it is made of blue light and red light. Color vision is in fact very complicated. We sometimes see what appears to be a spectral color when only light of other wavelengths enters our eyes. The analysis of light by locating maxima and minima in an interference pattern makes it possible



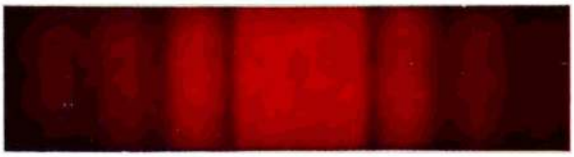
9–A Interference pattern produced by white light passing through two narrow slits.



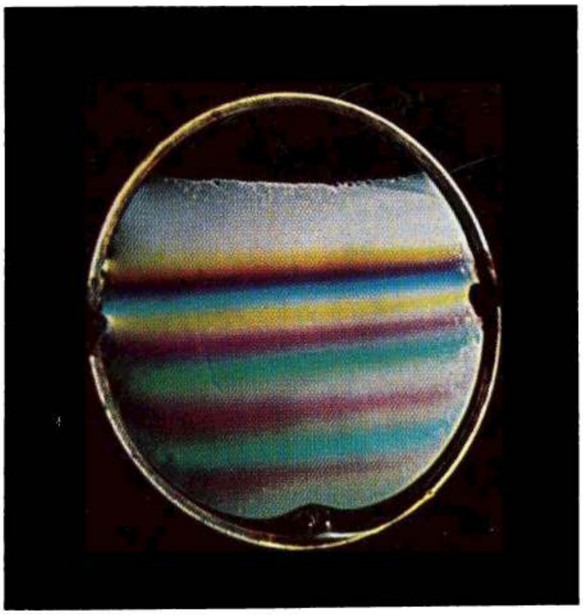
9–C White light passing through a single slit produced this pattern.



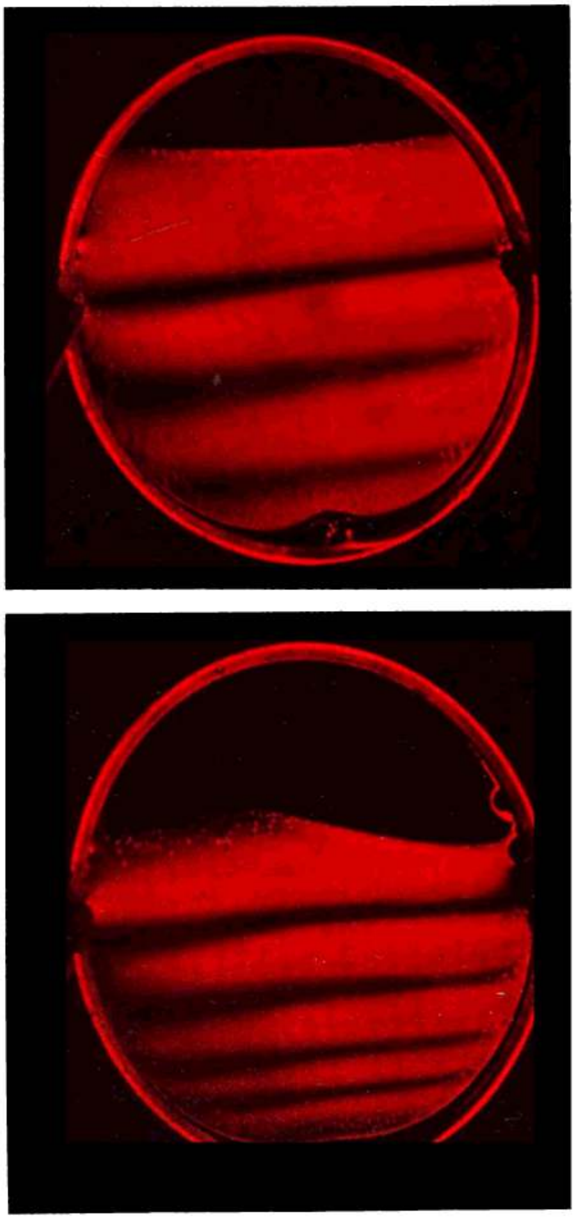
9–B Interference patterns of red light and blue-violet light made with exactly the same setup used to make the white light interference pattern in Fig. 9–A.



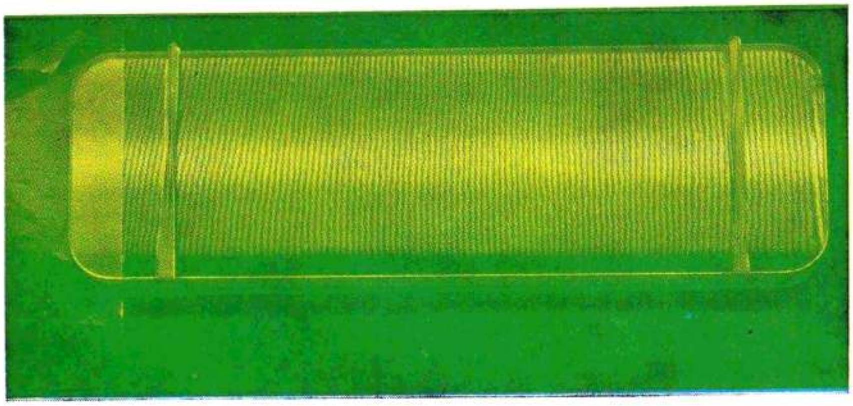
9–D With the same setup as in Fig. 9–C, red light produced this pattern. Both patterns are called diffraction patterns.



9–E Interference seen in white light reflected by a thin soap film. The film fills a circular aperture. The picture was taken shortly after the film had drained enough so that the upper region is less than $\lambda/4$ thick for all wavelengths of visible light.



9–F Right: interference seen in red light reflected from the same soap film. Top: at almost the same time that Fig. 9–E was photographed. Below: at a later time. Because the film has drained more, the dark region extends farther down at the top.

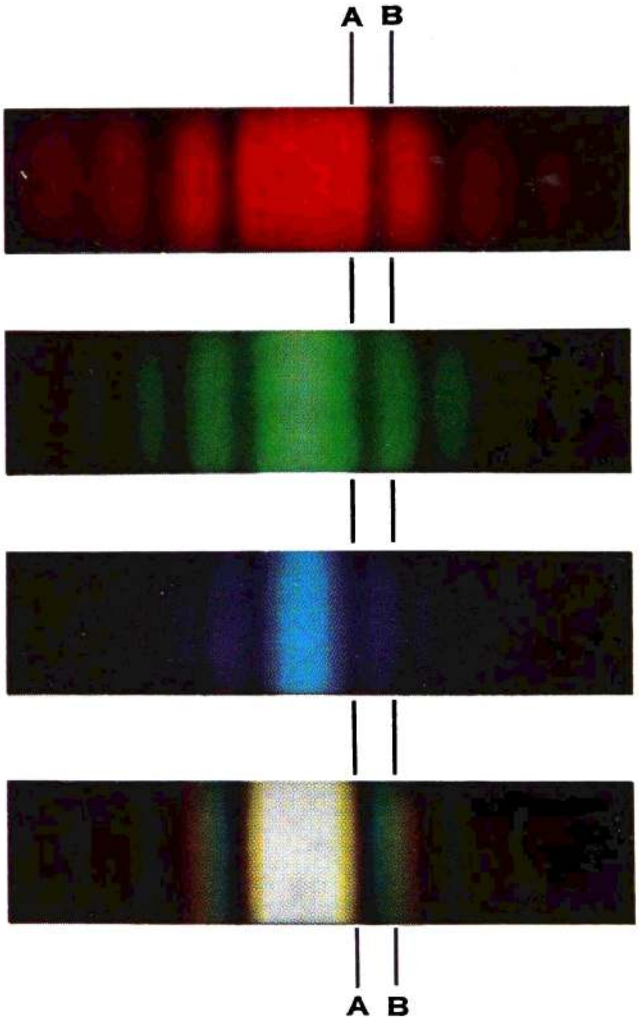


9–G **An interference pattern produced when light of one spectral color is reflected from the two sides of an air wedge made by separating two pieces of plate glass by a thin sheet inserted at the left end.**

How a Colored Diffraction Pattern Is Formed by White Light

From top to bottom in Fig. 9 – H you see the diffraction patterns of the same slit in red light, green light, blue light, and white light. If you look down the center line, you see that light of every spectral color is at a maximum. Their combination in the pattern at the bottom looks white. Now look down the line AA. The red light is moderately intense; the green almost absent; and the blue completely gone. In the white light pattern we see red. A little bit to the left of the line AA in addition to red, there will be more yellow and green and there you see yellow. To the right of the line the red disappears but the intensity of blue rises. What do you see in the white light pattern?

Try the line BB, a little farther out on the same side of the center. Here the red light is almost absent, the green light is bright and the blue light has almost disappeared. The resulting yellowish color in the white light pattern arises from the yellow-green region of the spectrum. Run down any line of your own choosing and see how the white light pattern is formed.



9–H **Formation of a white light diffraction pattern from the diffraction patterns of the spectral colors.**

to characterize the light accurately in terms of wavelength. Thus, interference patterns extend our knowledge far beyond what we can learn with unaided eyes.

The experiments we have been describing are ones that you can do yourself. Indeed, you can check up on our evaluation of the wavelengths of red and blue-violet light just by making some measurements of Fig. 9 – B. This photograph was taken with a single source of white light and two slits. We made the different patterns that you see, one above the other, by placing a red-colored filter above a blue-violet-colored filter in the horizontal path of the light between the source and the photographic film. The ratio of the spacing of the bars in the two halves of the picture should correspond to the ratio of the wavelengths of red and of blue-violet light. By measuring the photograph we find that the separation of the bars in the red pattern is greater than that in the blue-violet by a factor of about 1.4. We can compare this to the ratio of the wavelengths measured independently. That ratio is

$$\frac{\lambda_{\text{red}}}{\lambda_{\text{b-v}}} = \frac{6.5 \times 10^{-7}\text{ m}}{4.5 \times 10^{-7}\text{ m}} \approx 1.4$$

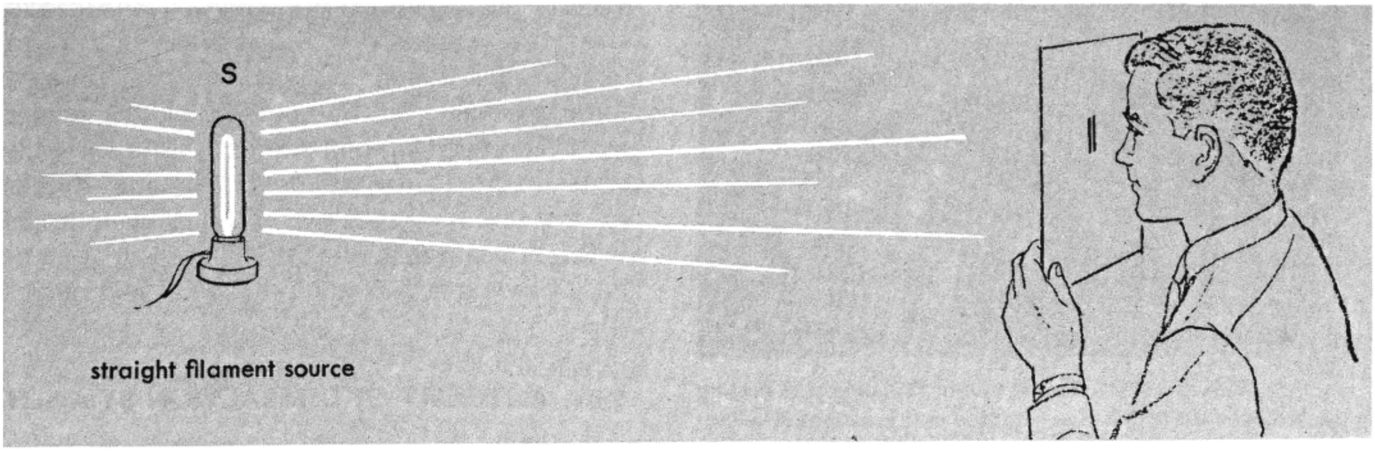
You can also carry out the complete set of experiments and measure the wavelength of light. Make your own slits and use your eye in place of the camera. You can cut the slits by pulling two razor blades side by side across a piece of glass covered with carbon. (See the Laboratory Guide.) The only other equipment you need is a strong narrow source of light and some filters of colored plastic or glass. For the light source, a showcase lamp in which you can see the straight filament is a good choice (Fig. 9 – 6).

We have now succeeded in showing that the spectral colors of light are related directly to the wavelength in vacuum. We have measured the wavelength of light of one color by measuring the spacing of the bars in the interference pattern from two slits. We therefore have a primitive spectroscope whose operation we understand.

A much more efficient spectroscope that sorts out the colors of light by interference can be made by using many equally spaced slits. The effect of the many slits is to intensify the amount of light emerging from the slits in any direction and to make the maxima for any wavelength sharper. These are known as grating spectroscopes and are commonly used to analyze light into its component wavelengths. For some purposes these spectroscopes are superior to those that use a prism to spread the light out into a spectrum.

9–5 DIFFRACTION: AN INTERFERENCE EFFECT IN SINGLE SLITS?

If you scratch only one slit with a razor blade (or a needle to make a slightly wider opening) and look through it at a white light source, you will see something like the photograph in Fig. 9 – C. There is a bright, broad central region surrounded by colored regions of lower intensity. In light of one color the pattern is similar. Figure 9 – D shows what you find if you use a red light source—a bright center, then dark re-

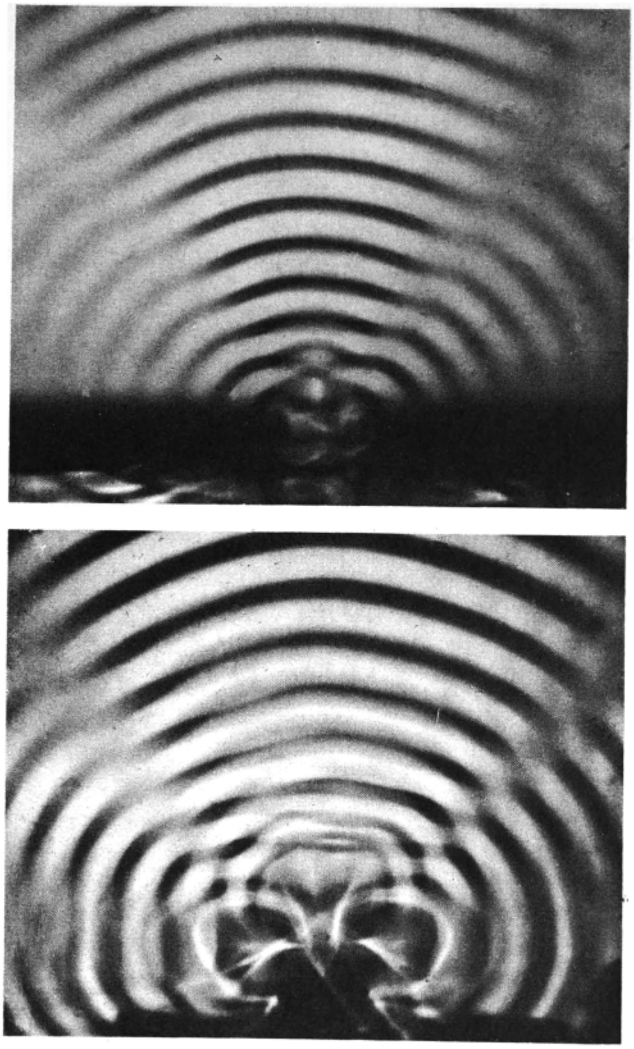


9–6 Observing a two-slit interference pattern. The lens of your eye brings the light from the two slits together to form an interference pattern on the retina of your eye.

gions alternating with progressively less intense regions of light.

Although the pattern of light and dark differs from Young's pattern, it looks suspiciously like an interference pattern. There are nodes and maxima in light of one color, and color effects when white light is employed. Because this interference effect takes place with only one slit, it may seem to make our interpretation of Young's pattern questionable. To clarify the situation we shall study more carefully the behavior of light passing through a single slit.

If we look back at our earlier discussion of the diffraction of waves by a slit (Section 9 – 7), we notice that the single slit does not act like a point source of waves unless its width is less than a wavelength. Normally, the slits we use in observing diffraction or interference in light are much wider than the wavelength of light. (Both the narrower slits we scratch with a razor blade and the wider ones made with a needle are many wavelengths wide.) Perhaps they are 10^{-1} mm wide while the wavelength of light is between 4 and 7×10^{-4} mm. The explanation of the interference patterns we see in light passing through single slits must be found by carefully examining the behavior of light as it passes through a narrow slit which is, nevertheless, many wavelengths wide.



9–7 Above, a diffraction pattern of straight waves passing through a slit and, below, an interference pattern of a line of equally spaced point sources extending across the slit. Near the sources the effect of source separation leads to some difference in the patterns. Far away the two patterns are the same.

In Fig. 7 – 19 we saw how straight waves traveling through a narrow slit in a ripple tank spread out or diffract. The waves on the far side of the barrier arise from the propagation of the parts of the crests and troughs that enter the slit. Instead of producing the crests and troughs in the slit by letting a straight wave fall on it, it seems possible that we might produce them by a line of point sources moving up and down together at the position of the slit. As a test of this idea let us replace the slit by a line of very small point sources, separated by a small fraction of a wavelength. We shall use enough sources so that they just fill the space of the original slit. If we now run the sources in phase and at the same frequency as that of the original wave, we see (Fig. 9 – 7) that the pattern produced does look like the diffraction pattern of the slit.

It seems reasonable that the wave patterns formed when straight waves are diffracted by a slit are the same as the patterns formed by the waves from a large number of point sources evenly spread through the width of the slit. We shall now assume this equivalence, and try to explain the interference patterns we observe through single slits by superposing the waves from the effective point sources in the slit. This general procedure of predicting the further propagation of waves by replacing wave fronts with sources was applied by Huygens to a number of problems including reflection and refraction.

9–6 A THEORY OF DIFFRACTION BY A SLIT

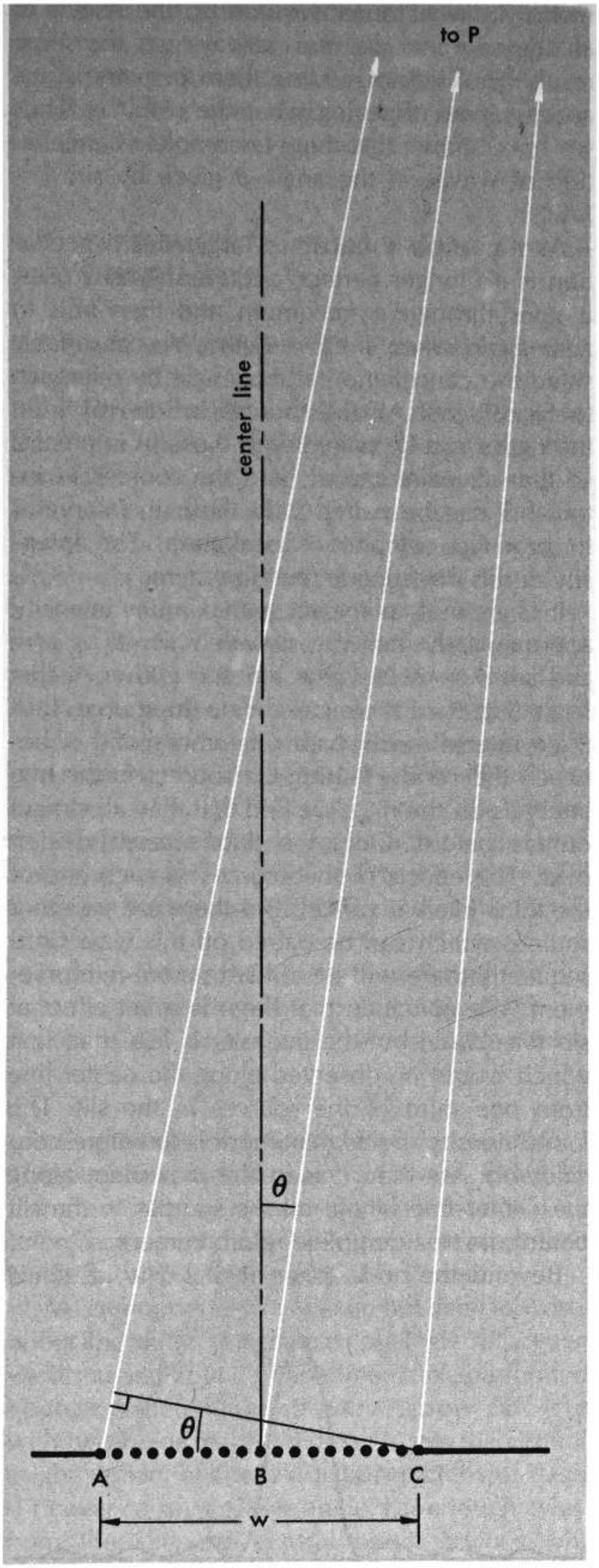
In our study of the interference pattern from a line of closely spaced point sources occupying the width of a slit, we shall consider the pattern only at distances very large compared to the slit width. Waves reaching such a distant point have traveled away from the sources along almost parallel paths (Fig. 9 – 8). We start with the total wave produced by the sources at a distant point directly in front of the slit. The path lengths from all the sources to that point are almost equal, and crests from all the sources arrive there together. This reinforcement of all the individual waves means that strong waves go out from the slit along the center line.

Next we examine a point *P* some distance off the center line so that the angle θ between *BP* and the center line in Fig. 9 – 8 is no longer zero. (Point *B* is at the center of the slit.) Because *PB* is essentially parallel to *PA*, the angle between *PA* and the center line is also θ. But *PA* is perpendicular to the wave front; the center line is perpendicular to the line of sources; hence, the angle between the wave front and the line of sources is θ, also. The path lengths from the various sources to point *P* are no longer equal. In particular, *PC* is shorter than *PA*. The individual crests do not arrive at *P* at the same time, and the total wave is weaker than it is along the center line.

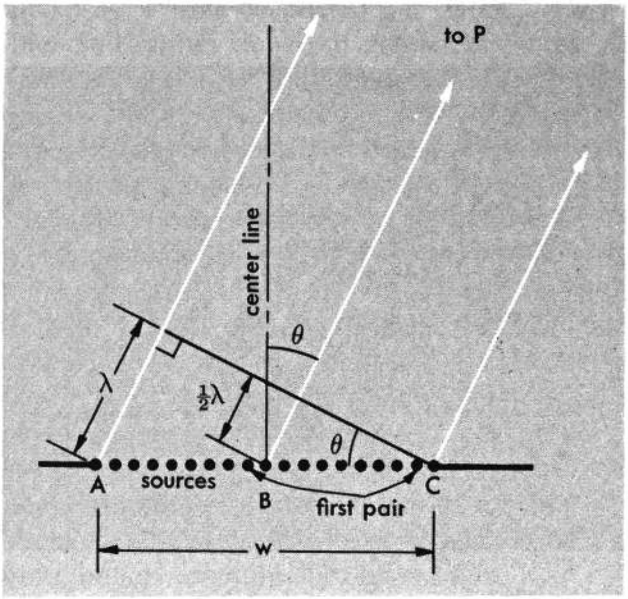
We now move *P* farther away from the center line so that the angle increases until the difference in the path lengths to the two ends of the slit becomes

$$PA - PC = \lambda$$

At this point $\sin\theta = \lambda/w$ as we see in Fig. 9 – 9. Also, as the figure shows, $PB - PC = \lambda/2$. We

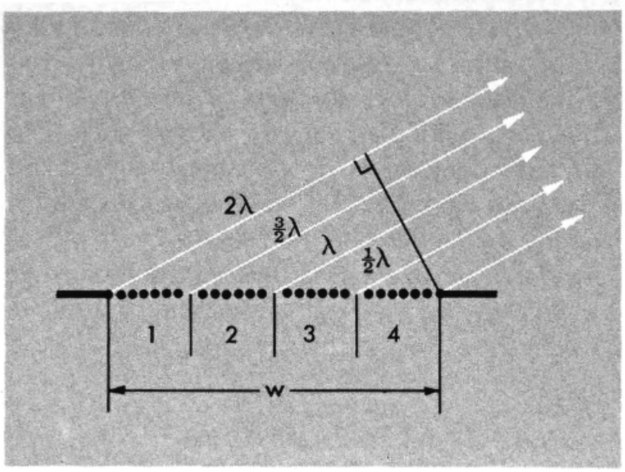


9–8 Light rays from the various sources to a distant point P (off the top of the page) are nearly parallel. However, if the distant point is not on the center line, the paths are of different lengths.

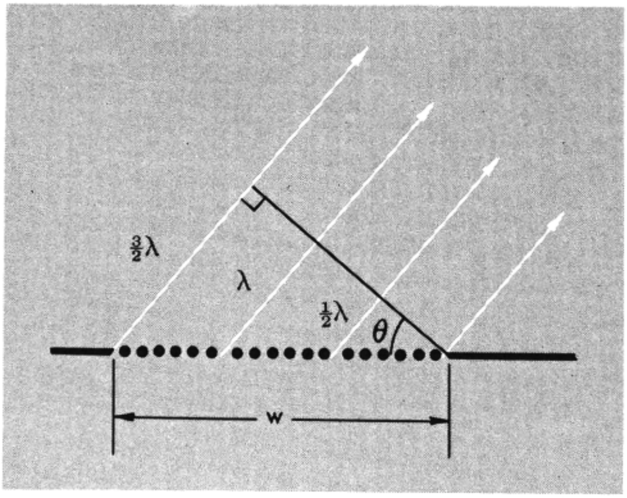


9–9 When $\sin\theta = \frac{\lambda}{w}$, waves from the sources cancel.

can now show that the waves from the individual sources reaching any faraway point *P* at this angle cancel each other so that the total intensity far away in this direction is zero. To understand this cancellation we match the sources in pairs which cancel. Consider the sources just to the left of *C* and of *B*. They give crests at *P* which come $\frac{1}{2}$ wavelength apart. In other words, at *P* a trough from one of these sources arrives with a crest from the other, and they cancel. Moving to the next pair of sources to the left, the second points to the left of *C* and *B* respectively, the same thing happens. Again we get the same result for the third pair, the fourth pair, and for all succeeding pairs. We have thus superposed the effects of all the sources; they all cancel in pairs; and no resultant disturbance takes place at *P*. It does not matter in what order we add up the effects of all the sources; we must always get the same result. The trick of adding them in pairs is just an easy way of seeing what the result is. Thus we have shown that there is complete cancellation of waves at the angle θ given by $\sin\theta = \lambda/w$.



9–10 When $\sin\theta = \frac{2\lambda}{w}$ there is cancellation.

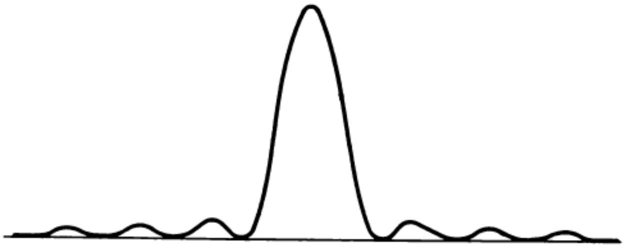


9–11 When $\sin\theta = \frac{3\lambda}{2w}$, partial reinforcement occurs.

As the angle θ becomes larger the cancellation is no longer perfect, and the intensity rises. It goes through a maximum and then falls to zero again when $\sin\theta = 2\lambda/w$. We can understand the cancellation at this angle by reference to Fig. 9 – 10. All the sources in interval 1 on the figure can be paired with those in interval 2 so that all pairs cancel; and the sources in interval 3 can be paired with those in interval 4 to produce complete cancellation. The intensity in this direction is therefore zero.

It is natural to expect a maximum intensity approximately halfway between $\sin\theta = \lambda/w$ and $\sin\theta = 2\lambda/w$, near $\sin\theta = \frac{3}{2}\lambda/w$. At this angle (Fig. 9 – 11) we can divide the sources into three intervals with path differences of $\frac{1}{2}\lambda$ between their ends. Pairing the sources in the two intervals on the right we find that they all cancel out as before. Only the third interval is left over. The effects of the sources at each end of the third interval cancel. But there are no other sources which can be paired off this way. Consequently there will be at least partial reinforcement. We conclude that there is a net effect at $\sin\theta = \frac{3}{2}\lambda/w$; but the intensity is less than that which would be observed along the center line from one third of the sources in the slit. The light intensity in this maximum is therefore considerably less than it is in the maximum along the center line where all the sources in the slit contribute to a complete reinforcement.

Beyond the node at $\sin\theta = 2\lambda/w$, as $\sin\theta$



9–12 A single-slit diffraction pattern for light of one definite wavelength. The intensity of the light is plotted vertically as a function of the distance from the center line.

increases, the intensity rises to a still weaker maximum near $\frac{5}{2}\ \lambda/w$, and falls to zero again at $3\ \lambda/w$. It continues its rise and fall as θ increases, while the maxima get progressively weaker, as shown in Fig. 9 – 12.

9–7 EXPERIMENTAL CHECKS WITH SINGLE AND DOUBLE SLITS

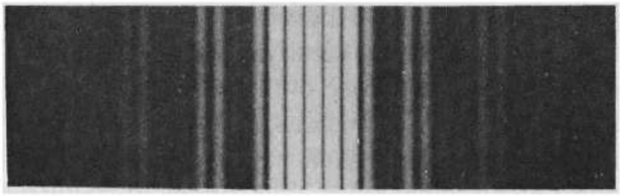
We now have a theory which connects the appearance of the diffraction pattern of a slit with the width of the slit and the wavelength of light. Does the theory account for actual observations? To be sure, we must construct slits of accurately known width and measure the distances from the center line to the dark regions which occur in the diffraction pattern of light of a single color. When we make such measurements, we find that the distance between the nodes at the sides of the pattern is constant. Also, as expected, the distance across the central bright band is twice the distance between the other nodes. (Measure it yourself on Fig. 9 – D opposite page 132.) Furthermore, the intensities of the maxima decrease with the distance from the center, as predicted by the theory. We therefore have reason to believe that the theory is good.

Finally, we can find the wavelength of the light by measuring the distance between the center of the pattern and the first node. Using

$$x/L = \sin\theta = \lambda/w$$

we find values of λ for light of various colors. These values are the same as those found by using two slits or a many-slit spectroscope. Our interpretation of both Young's experiment and single-slit diffraction is probably correct.

We can now see why we did not need to worry about the diffraction pattern of single slits when we discussed Young's experiment. We performed Young's experiment with very narrow slits (cut with two razor blades). For such slits the central bright region of the diffraction pattern of each slit covers a rather large angle. Far enough away from the slits the central bright regions will largely overlap, and in this overlapping region we see the simple double-slit pattern. The other maxima of the single-slit diffraction pattern occur far out to the sides, and they are too weak to be seen easily. For this, the slits have to be wider. (See Fig. 9 – 13.)



9–13 When two identical, fairly wide slits are used to make an interference pattern, and when they are separated by a distance comparable to their width, the resulting pattern combines the features of a single-slit diffraction pattern and the two-slit interference pattern. (Courtesy: Bruno Rossi, "Optics," Addison-Wesley Publishing Co., 1957.)

9–8 INTERFERENCE IN THIN FILMS

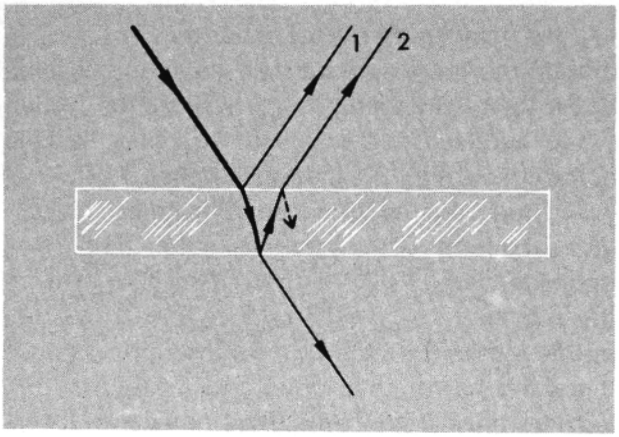
Of the many interference patterns we can see, among the most common are those we observe when light is incident on a thin soap film. When white light falls on a thin film it gives striking color effects (Fig. 9 – E opposite page 132).

To study this effect we shall examine the pattern obtained with a single color (Fig. 9 – F opposite page 132). In the reflected light from the film there is a wide dark region at the top. In this region the soap film is very thin, because the water has drained down toward the bottom, leaving only a few clinging molecular layers behind. Farther down the film, we notice that the thickness increases, until finally we come to a thickness at which reflected light is clearly visible. Still farther down, at greater thickness, a dark bar appears in the photograph; the reflection is absent. Then continuing farther, bright and dark bars succeed one another as the thickness gets still greater.

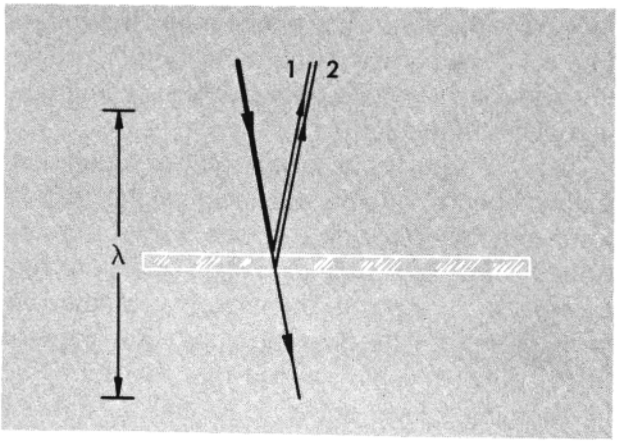
That the observation of a bright or a dark bar really depends on the thickness of the film is borne out by watching the bars while a newly formed film is draining. When the top has drained to a certain thickness, a bright bar forms there. As the draining continues, the bar moves down the film, and a dark bar forms in the thinning region above. Bar after bar moves down the film, each staying with its appropriate thickness. Finally a dark region spreads down from the top when the film is very thin.

Experiments with wedges—air between plate glass, for instance—show the dependence on thickness even more clearly (Fig. 9 – G opposite page 133). As the angle of the wedge is reduced, the bars, which in this case are equally spaced, move apart so that each bar stays exactly at its appropriate thickness in the wedge.

How can we explain these alternating bright and dark bars? They are very similar to the alternating bars in Young's interference pattern, and their explanation is almost the same. When light passes through a sheet of refractive material, some of the light is reflected at the first surface it meets, and some is reflected at the second surface. Take a look at Fig. 4 – 15, or think of your reflection in a glass window when you sit in a lighted room on a dark night. If the glass is thick, the two reflections are sufficiently far apart to be apparent. Of course, a small part of the light that has been reflected at the bottom of the glass will be reflected again at the top instead of passing out. Some light may be partially reflected several times inside the glass, but for a weakly reflecting material like glass or a soap film, only the reflection as the light enters and the first reflection inside are usually appreciable (Fig. 9 – 14).



9–14 Rays 1 and 2 represent the two most important reflections from a thin film.



9–15 Although there is negligible path difference between beam No. 1 reflected from the top of a very thin film and beam No. 2 reflected from the bottom, the two reflected beams cancel.

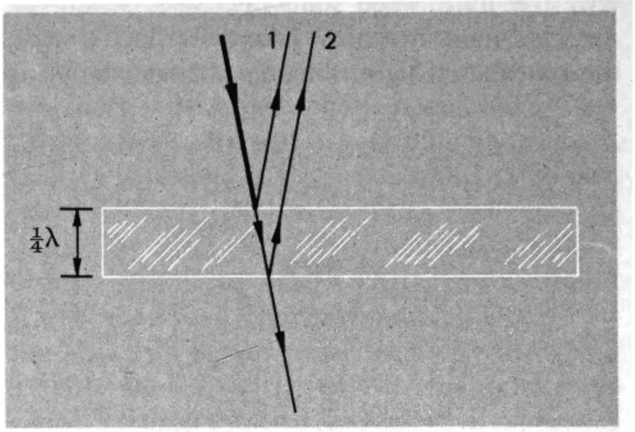
For thin soap films, as well as for thick glass, the two strongest reflected beams of light travel over different paths. We shall call the beam which is reflected at the outside surface No. 1, and the beam which has passed through the film and has been reflected at the other side beam No. 2. It is the interference between these two beams that produces the bright and dark bars in the reflected light of a single color and the color effects when white light is incident.

To understand the interference patterns most easily, let us consider only the case of light striking the film normal to the surface. In the figures we shall refer to (Fig. 9 – 15 through 9 – 19), we have drawn the rays tipped a little from the normal in order to be able to tell them apart. (Of course, these "normal rays" would not show any bending as they go through the surfaces.) We shall start with a very thin film like the one at the top of the soap film shown in Fig. 9 – F. As you know from experiments with monolayers, even if the film in this region is many molecular layers thick, it is still very thin compared with the wavelength of visible light. For yellow light the wavelength is about 6×10^{-7} meter, while the thickness of a film 100 atoms thick is only 10^{-8} meter. Our thin film is illustrated in Fig. 9 – 15. Because its thickness is much less than the wavelength of the light, the path difference between the two beams is very small compared to a wavelength, and we might expect that the two reflected beams add together so that a reflected crest in one and a reflected crest in the other are almost superposed. The reflected light should then be appreciable. Instead, we see that no net reflection occurs. The whole top of the soap film is dark. Apparently the waves in the two reflected beams cancel each other instead of reinforcing.

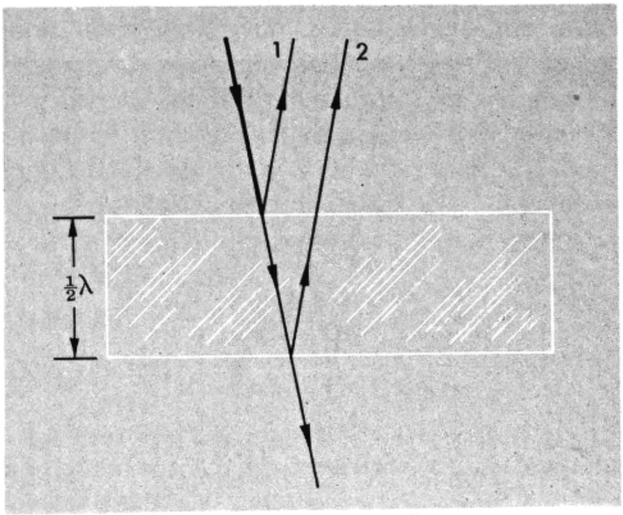
This cancellation may seem puzzling; but when we recall the behavior of waves reflected on coil springs, it is easier to understand. When a pulse traveling along a spring is reflected from a heavier spring in which the speed of waves is smaller, it turns upside down. When a pulse is reflected from a lighter spring where the speed is greater, it stays right-side up (Section 6 – 4). Here with light we have an analogous situation. The light beam reflected from the outside of a soap film is reflected from a more refractive material in which the wave moves more slowly, while the light beam reflected inside the thin film is reflected from air, a less refractive medium in which the wave moves faster. Perhaps we can picture the more refractive medium as acting like a heavy spring. Then light waves reflected from the outside of the film into beam No. 1 should turn upside down. An incident crest turns into a reflected trough. On the other hand, an incident crest reflected inside the film will be reflected into beam No. 2 as a crest. Because one wave is turned upside down and the other is not, the two waves cancel; and we see no reflected light.

Let us now consider what happens as the soap film becomes thicker. Beam No. 2, which travels back and forth through the soap film, has a longer path than the beam reflected from the top surface. In beam No. 2, therefore, the crests fall farther and farther behind as the thickness is increased. When the thickness becomes one quarter of a wavelength, the path in the soap film is a half wavelength long; and the crests in this beam have dropped behind by a half wavelength (Fig. 9 – 16). Instead of coinciding with troughs in beam No. 1, as they do for a very thin film, they have now been delayed until they coincide with crests. The crests in the two beams add together, and the total reflected light intensity is at a maximum.*

As we increase the thickness of the soap film still further, the crests in the internally reflected beam fall still farther behind. When the film is half a wavelength thick, the length of the path back and forth inside the film is a full wavelength. Since the crests in beam No. 2 have fallen behind by a whole wavelength, they now coincide with troughs in beam No. 1 just as they did when the film was extremely thin (Fig. 9 – 17). The crests and troughs in the two reflected beams cancel again, and we see a dark bar in the light reflected from the soap film at this thickness. As we continue to increase the thickness, every quarter wavelength we shift from a region in which there is no reflection to a region of maximum reflection, or from a region of maximum reflection to a region of no reflection again. There is no reflection from the film when its thickness is $\frac{1}{2}\lambda$, λ, $\frac{3}{2}\lambda$, etc. Maximum reflection occurs when the film is $\frac{1}{4}\lambda$ thick, $\frac{3}{4}\lambda$, $\frac{5}{4}\lambda$, and so on. It is for this reason that the bright and dark bars seen in a thin wedge are equally spaced. The thickness of the wedge increases at a constant rate, so the bars alternate with constant spacing.



9–16 Light reflected from the top and bottom of a film that is $\lambda/4$ thick. Because of the extra path length $\lambda/2$ in beam No. 2, and because the waves in beam No. 1 are turned upside down on reflection, the two reflected rays reinforce.



9–17 The two reflected rays when the film is half a wavelength thick. The path difference is λ, and beam No. 1 is inverted on reflection. They cancel.

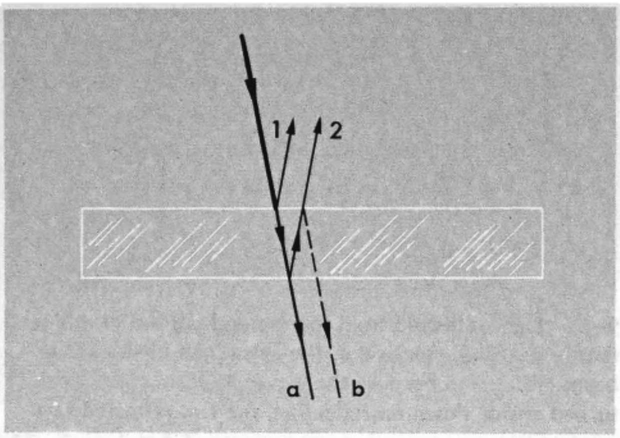
---

*Clearly, the wavelength we are using is the wavelength in the soap film.

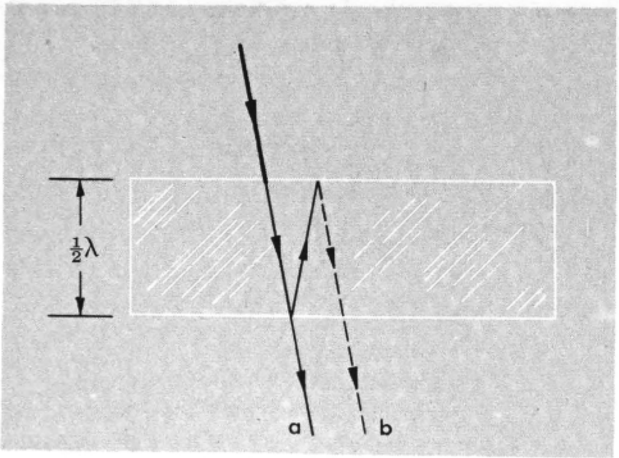
9–9 INTERFERENCE IN LIGHT TRANSMITTED THROUGH THIN FILMS

When the intensity of reflected light changes as a function of thickness, what happens to the light transmitted by the film? When there is no reflected light we expect that the intensity of the transmitted light will be the same as that of the incident light. Experiments agree with our expectation. At thicknesses $\frac{1}{2}\lambda$, λ, $\frac{3}{2}\lambda$, etc, all the light is transmitted. Furthermore, at those thicknesses for which reflection takes place, the intensity of the transmitted light is less by just the intensity of the reflected light.

We can understand in terms of interference why the transmitted light decreases when the reflected light increases. Like the reflected light, the transmitted light is made of two interfering beams. They are the beam of light that passes straight through the film without any reflections and a beam made of the light which is internally reflected twice (Fig. 9 – 18). Other internal reflections give beams too weak to be important.



9–18 The unreflected ray, a, and the ray that is twice reflected, b, are superposed in the transmitted light.



9–19 The transmitted ray and the twice-reflected ray reinforce when the film is λ/2 thick. The extra distance traveled by b is λ.

We now examine how the two beams *a* and *b* in Fig. 9 – 18 add up for two different thicknesses of the film, a thickness at which there is no reflection and a thickness at which there is maximum reflection. There is no reflection when the film is $\frac{1}{2}\lambda$ thick. In this case (Fig. 9 – 19) the twice-reflected wave *b* is delayed by a whole wavelength with respect to the unreflected wave *a* because of its extra journey back and forth across the film. A crest in the twice-reflected wave, therefore, comes out of the bottom of the film along with a crest in the unreflected wave. The two crests must be added together, and the net effect is to produce crests in the transmitted light which are larger than those in the unreflected beam.

On the other hand, when there is a maximum reflection, at the thickness $\frac{1}{4}\lambda$ for example, the twice-reflected wave *b* travels an extra distance $\frac{1}{2}\lambda$ compared to the unreflected beam *a*. The crest in *b* therefore coincides with a trough in *a* and the superposition of *a* and *b* results in smaller troughs and smaller crests in the transmitted light. We thus see that when there is no reflected light from the film, the transmitted beam is more intense; and when there is a maximum in the reflection from the film, the transmitted light is less intense. If we work through the details quantitatively, we find that the sum of the intensity transmitted and the intensity reflected is, indeed, equal to the intensity of the incident light.

9–10 COLOR EFFECTS IN INTERFERENCE

In explaining interference patterns we have concentrated on those produced by light of a single spectral color. We can now go back to the interference patterns of white light, shown opposite pages 132 and 133, and explain the many bright colors in those patterns. We consider the interference pattern of each spectral color separately; then, by noting which spectral colors are present and which are absent at a given place, we can predict the color that we shall see at that place in the white light pattern. For example, if at one place blue light is absent but red light is intense, we shall see a color belonging to the red end of the spectrum—yellow or red. If at another place only blue and red light are present, we shall see purple. The procedure is illustrated in Fig. 9–H and in the accompanying text opposite page 133. There it is applied to the white light diffraction pattern of a single slit. You can apply the same procedure to Young's interference pattern, to thin film interference patterns, and in fact to any other interference pattern.

9–11 CONCLUSION

In this part of the course we have examined the behavior of light. We started with its simplest properties, propagation through vacuum, reflection, and refraction. These we tried to explain with a particle model. We have gone on to the characteristic wave behavior of light, examining diffraction and interference. With the aid of the ripple tank, we have seen this kind of wave behavior directly.

We now see that light behaves like waves. Not only does it reflect and refract like waves; it also exhibits interference effects. Reflection and refraction may be explained on a simple particle theory, but interference is the identifying characteristic of waves. Since the demonstration of interference in light, the wave nature of light has never been doubted.

Although light is an old interest, new aspects are still being discovered. Just in our century, we have shown that X rays are light waves, and we have measured their wavelengths by observing interference patterns. But we have also discovered that light exhibits some particle-like behavior. In certain circumstances we can even count these particles, as we shall see later on. We now are beginning to understand how to describe both the wave and particle aspects of light in one theory.

You probably have many questions to ask about light. For example, just how is the measured intensity related to the height of the waves? We shall not answer this question here, but you can learn about it. Probably most of the questions you will ask have intrigued other people, and something is already known about them. A number of them have given rise to startling developments. Take one example: through what medium is light propagated? Its propagation is influenced by various materials, but it does pass through vacuum. If we assume the existence of a medium for the propagation of light through vacuum (it has been named "the ether"), we later find that this ether is impossible to detect. The failure of measurements designed to show the presence of the ether was one of the important puzzles that led Einstein to formulate his theory of relativity.

We can study more about light, and we can also study more about waves in general. The relation of wave height to intensity is really a general question. So are many others. To illustrate we shall take a last example.

You may have noticed that some rooms (usually those with large, unbroken wall areas) have "dead spots" where it is very difficult to hear a note of a certain pitch. Nearby, this same note may be clearly audible. You can easily demonstrate this effect with a high-pitched whistle. It occurs because of interference between direct waves from the source and waves reflected from the walls. At some frequencies the nodes and maxima form a fixed pattern in the room, commonly referred to as a standing wave pattern.

While using the ripple tank you have undoubtedly seen waves hitting a reflecting barrier head on and traveling back in the opposite direction. If the incident and reflected waves are of about equal height, close to the barrier you can see regularly spaced regions of no motion alternating with zones where the water appears to move up and down. In this region we do not see the traveling waves propagating in either direction. Here again we have an example of a standing wave interference pattern. Standing wave patterns occur for all kinds of waves. We shall return to them briefly when we study matter waves.

QUESTIONS

1.* In a ripple tank the size of a lake, with long waves on it, how would you search for nodal lines with a boat in the dark?

2. Why can't we see interference from two incandescent filaments?

3. In an interference pattern produced by white light passing through two narrow slits, the distance between the black bars is 0.32 cm. The distance between slits is about .02 cm, and the distance to the screen on which the bars are observed is 130 cm. Find the average wavelength of white light.

4. A source of red light produces interference through two narrow slits spaced a distance $d = .01$ cm apart. At what distance from the slits should we place a screen so that the first few interference bars are spaced one centimeter apart? What will be the spacing of the bars if we then use violet light?

5. When a source of light of wavelength λ is used in a two-slit experiment with narrow slits separated a distance d, at what angles do you expect to find the *maxima* in the light intensity in the interference pattern?

6. A double-slit source of constant phase delay can be produced, using the direct light from a single slit and the reflected light from a glass plate, as shown in Fig. 9 – 20. The interference fringes can be observed by means of a simple magnifier focused on the edge of the glass plate. When observed in this manner, the dark lines are found to be equally spaced, starting at the reflecting surface of the plate.

   (a) What does this indicate about the phase delay between the two sources?

   (b) If the wavelength is 5400 angstroms and the black bars are 0.9 mm apart, how far is the source slit above the plane of the surface of the glass plate when the slit is 60 cm from the edge of the glass plate?

7.* How far would light travel in air in 3×10^{-9} sec?

8. Calculate the period of yellow light. About how many wavelengths are included in a light wave during emission of a light burst by a single atom?

9.* A double-slit interference pattern was produced with the yellow light of sodium, and a spacing of 1.0 mm between nodes was obtained. What would be the spacing if violet light of potassium were used in the same system?

10.* What is the ratio of the spacing of the nodes in Fig. 9 – H (facing page 133) between red and blue light?

11. Suggest an optical method for measuring the width of a narrow slit.

12. What happens to the diffraction pattern from a single slit when you turn the slit as in Fig. 9 – 21 instead of holding the opening perpendicular to the path of the light? Why?

13.* Why is the partial reinforcement referred to in Fig. 9 – 11 for a single-slit interference pattern so much less than the central maximum for the same slit?

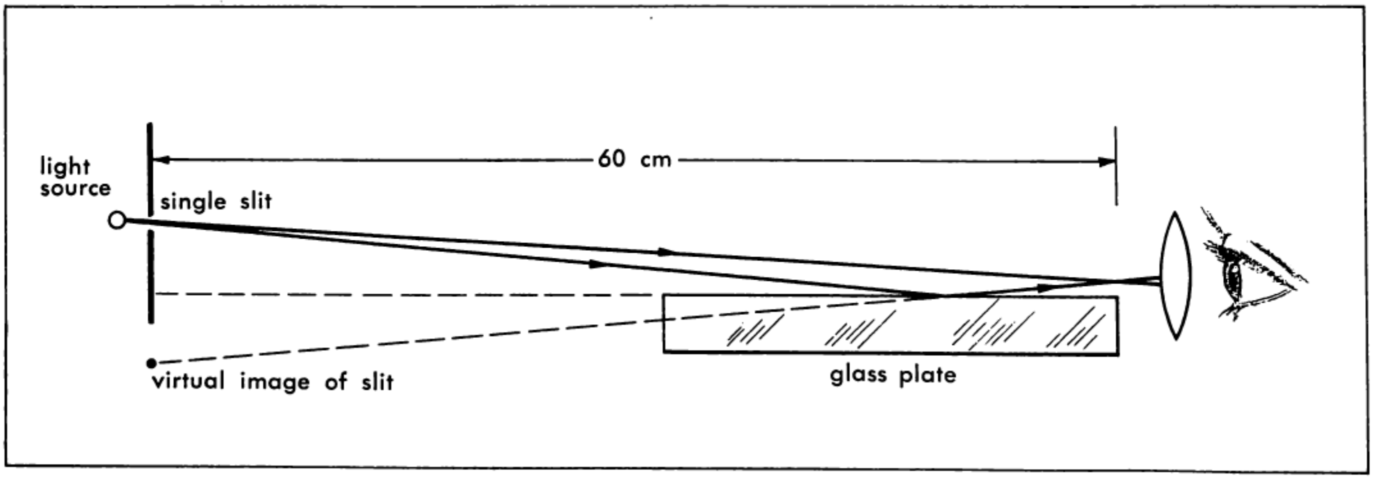
14. (a) When yellow light passes through a slit 1 mm wide, at what angles are the first three nodes in the diffraction pattern?

    (b) If the slit is 10 times as wide?

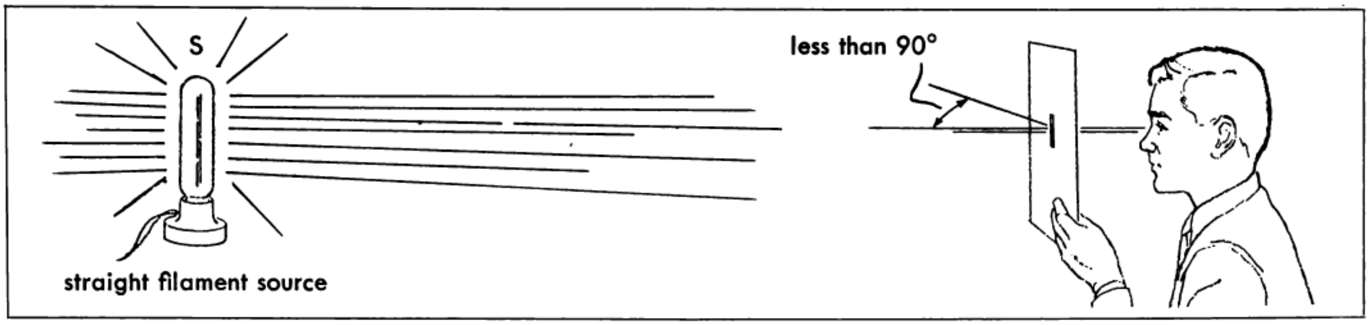
    (c) If it is $\frac{1}{10}$ as wide?

15. In Fig. 9 – 8, how will the diffraction pattern be affected

    (a) if you cover the right-hand half of the sources?



9–20 For Question 6.



9–21 For Question 12. The light rays pass obliquely through a single slit.

(b) if you cover the extreme left quarter and the extreme right quarter?

16. When a source of wavelength λ is used with three slits, each separated from its neighbor by the distance *d*, show that you get maximum intensity at the same angles as for two slits only.

17. Two 3-inch loudspeakers emit sound of a steady pitch whose frequency is 1000 vibrations per second. These sources are in phase and are 2 meters apart.
(a) At what angles would you expect to hear no sound? (The speed of sound is about 300 m/sec)
(b) What do you think would happen if you tried this experiment in a room with hard-surfaced walls?

18. In Fig. 9 – 13, identify the dark regions arising from the diffraction pattern of each slit and the dark bars arising from interference between the slits.

19.* What would happen to the distance between the lines in Fig. 9 – G (facing page 133) if red light were used instead of green?

20.* What is the wavelength in water of green light whose wavelength in vacuum is 5600 Å? (The speed of light in water is $\frac{3}{4}$c.)

21.* Several different materials are used to make $\lambda/4$ thick films for red light at 6100 Å. Will these films all have the same thickness?

22.* In Fig. 9 – G (facing page 133), if at the left-hand side you put 2 sheets between the glass plates instead of one, how will it affect the pattern?

23. From the photograph of a soap film at the top in Fig. 9 – F, plot a graph of film thickness against vertical distance down the film.

24. Two pieces of plate glass 10 cm long make an air wedge like that shown in Fig. 9 – G. The plates are separated at one end by a human hair with a diameter of 0.09 mm. The reflected interference pattern is observed by looking in a direction perpendicular to the surfaces of the plates.
(a) What is the spacing of the bars if the incident light is blue?
(b) How many bright bars are seen per centimeter if the incident light is red?
(c) Is the light reflected from the end where the plates are in contact a maximum or a minimum?
(d) Can you tell from this experiment which reflected waves are turned upside down?

25. In Fig. 9 – 21, (a) what is the path difference between rays *a* and *b*?
(b) How many times are the waves inverted?
(c) What is the overall phase delay between rays *a* and *b*?

26. Lenses are often coated with a thin film to reduce the intensity of reflected light.
(a) If the index of refraction of the coating is 1.3, what is the smallest thickness that will give a minimum reflection of yellow light?
(b) Such lenses often show a faint purple color by reflected light. Why?

27. Can you suggest reasons why no interference is seen when light reflects from the two surfaces of a windowpane?

10 Applied Optics

IN THE PRECEDING chapters we have looked for the fundamental laws which best sum up the behavior of light. To find these laws we experimented with light under the simplest conditions: we observed the reflection of light from a plane mirror, we studied the refraction of light at the boundary between two media, and we have seen the interference of light going through a double slit. The interplay of experiment and theory enabled us to predict the results of new experiments, but all our experiments were motivated by the desire to better understand the behavior of light.

In this chapter we wish to use our understanding of the behavior of light to construct useful instruments. We shall see that the knowledge of the laws of reflection and refraction and the understanding of the wave nature of light enable us to design devices such as telescopes and microscopes which will perform highly specialized tasks.

10–1 PARABOLIC MIRRORS

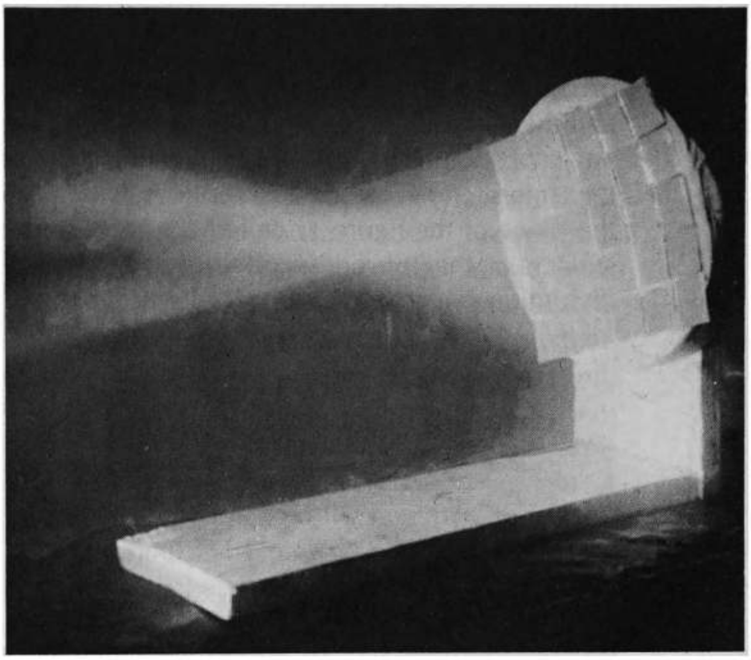
On the top of the Pic du Midi in the French Pyrenees research workers curious about the effects of ultrahigh temperatures have been able to use mirrors to concentrate enough sunlight to "cook" steel. In India, using inexpensive "mirror ovens," housewives cook dinner with concentrated sunlight. Astronomers use the large mirrors in their telescopes to concentrate the faint light from distant stars; they can thus produce on photographic plates images of stars that cannot be seen by the eye alone. What shape of mirror has this ability to concentrate light?

We cannot concentrate light with a plane mirror, for the light always diverges, appearing to come from the virtual image behind the mirror. But by using several plane mirrors, we can cause several pencils of light from the same source to cross each other in a small region of space. To be specific, let us assume that the source is far away—a star, perhaps—and therefore the pencils of light that come from it to our mirrors are almost parallel. The farther the source, the smaller the angles between the pencils of light, until for a star all the pencils reaching the earth are practically parallel. In Fig. 10 – 1 twenty-five mirrors are cocked at slightly different angles so as to reflect parallel light so that it passes through a single small region. We see the region because the smoke there scatters some of the light in the many crossing pencils. Other regions are equally smoke-filled but appear darker because few pencils of light cross there. The pencils of light pass through the small region of overlap and then diverge from one another, each going on just as it does when others do not cut across it.

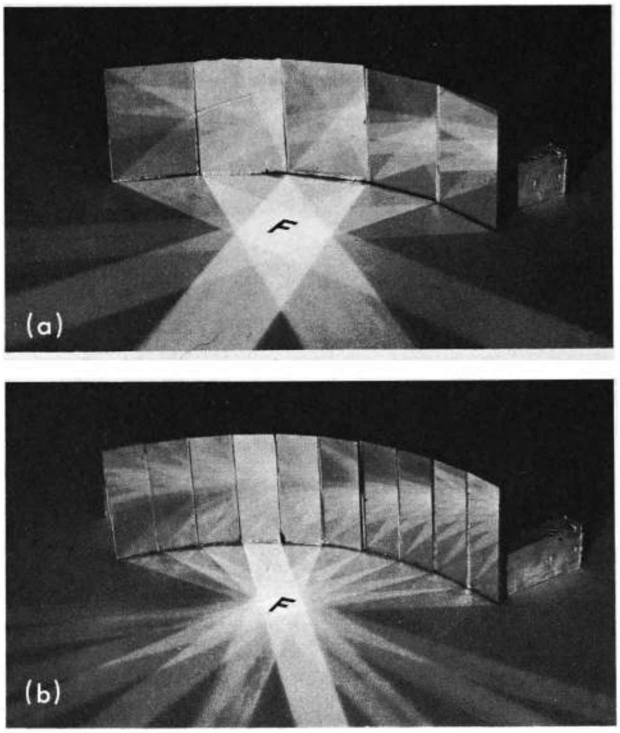
We can concentrate the light further if we use more and smaller mirrors. Let us imagine that we cut each one of our twenty-five mirrors into four pieces, for example. Then each pencil has only one-fourth of the previous cross-sectional area, and we can make them converge into a smaller region. Now we cut the mirrors again, making them still smaller. The region of overlap is cut down still more. If we keep doing this, eventually the mirrors are as small as we can make them; but they intercept the same amount of light that our original twenty-five mirrors did. They reflect the same amount and send it all through a single tiny spot.

To carry out the construction of a mirror which focuses a large amount of light at a single spot by successively dividing and adjusting the orientation of plane mirrors hung up in space is easy in concept but difficult in practice. We can illustrate the procedure, however, with the somewhat easier job of concentrating parallel light on a line instead of a spot. Two stages in this process which you can carry out without too much difficulty are shown in Fig. 10 – 2. It is easy to imagine the similar processes carried out in space to produce the single spot.

As we subdivide and reorient plane mirrors in space, we construct closer and closer approximations to a mirror with a continuously curved surface; and we can imagine the indefinite continuation of the process which results in a smooth mirror that focuses the parallel light. The shape of the smooth surface (called a *paraboloid of revolution* or *parabolic mirror*) is determined exactly by our imaginary procedure. The spot F to which all the reflected light converges is called the *principal focus* of the parabolic mirror. In Fig. 10 – 3 a cross section is shown through the principal focus parallel to the direction of the

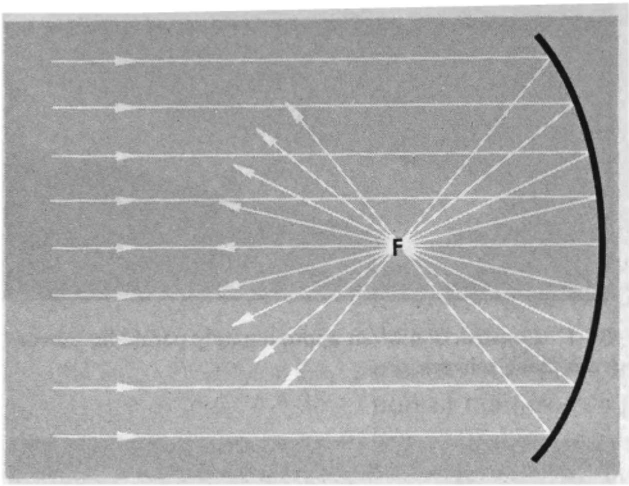


10–1 Construction of an approximately parabolic mirror from plane mirrors.

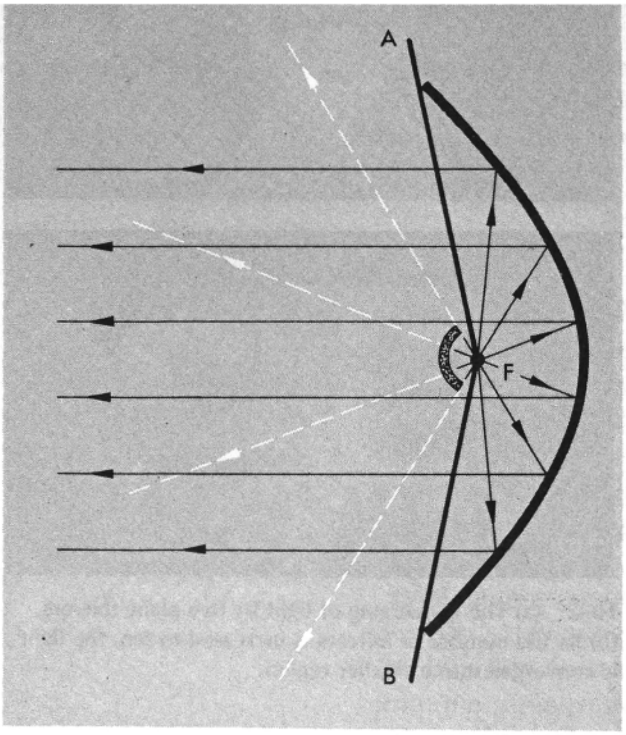


10–2 (a) The converging of light by five plane mirrors. (b) As the number of mirrors is increased to ten, the light is converged into a smaller region.

incident light. Each of the rays we draw indicates how the light is reflected. It stays in the plane of the figure according to the laws of reflection; and the angle of reflection of each ray equals the angle of incidence. The curve shown, which is the intersection of the parabolic mirror with the plane of the figure, is called a *parabola*. The surface of the mirror is called a paraboloid of revolution because it can be generated by rotating the parabola of Fig. 10 – 3 about the incident ray through the principal focus *F*.



10–3 The converging of light by a curved mirror.



10–4 A searchlight. The "black" rays from the source are reflected parallel; the white rays spread out unless the small "dark-gray" mirror reflects them back to the parabolic mirror.

When we make parabolic mirrors, we usually start by making a smooth surface that is approximately parabolic; then, if necessary, we improve it by distorting, grinding, and polishing until light sent in parallel to the axis of revolution is brought to an accurate focus.

10–2 SEARCHLIGHTS

We have seen that all of the light that falls on a parabolic mirror in a direction parallel to the axis is reflected in such a way that it passes through the principal focus. Since light can travel in either direction over a given path, we can, by putting a tiny intense source of light at the principal focus of a parabolic mirror, interchange the incident and reflected rays. It follows, then, that any light starting from the principal focus will, after reflection, be traveling parallel to the axis of the mirror. Many searchlights are constructed on this principle as shown in Fig. 10 – 4. All light from the source that strikes the parabola travels outward in parallel paths to form a narrow, intense beam that penetrates to great distances through space. The light emitted from the source within the cone *AFB*, however, never strikes the mirror, and it therefore spreads out. But even this light can be used if a small spherical mirror, shown dark gray in the figure, is mounted in front of the source in such a location that the center of the sphere is at the source. Such a spherical mirror will reflect all rays back through the principal focus because the rays will all hit perpendicular to the mirror and will be reflected back along the same path to the parabolic mirror. Of course, the small mirror will block off the light reflected from the central section of the big one, but this loss is not very important if the little mirror is made very small and is mounted close to the source at *F*.

How narrow is the parallel beam of light sent out by a searchlight? That depends on the opening of the parabolic mirror. Furthermore, the construction of the mirror also depends on the distance between the principal focus and the central point on the surface of the parabolic mirror. Once this distance is chosen, the shape of the parabola is fixed. The distance from the focus to the center of the mirror is therefore very important. It is given a special name and called the *focal distance* or the *focal length* of the parabola. We shall use the letter *f* for the focal length of any parabolic mirror.

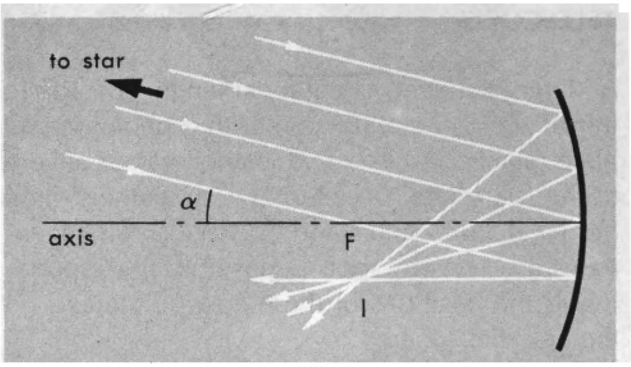
10–3 ASTRONOMICAL TELESCOPES

We know that a parabolic mirror brings to a focus all the light that arrives parallel to the axis of revolution; also, when we place a tiny source at the focus of such a mirror it sends out a parallel beam. What happens to light that does not come in parallel to the axis, or to light that diverges from a source which is not at the principal focus? The answers to these questions will lead us to understand such varied devices as astronomical telescopes and shaving mirrors.

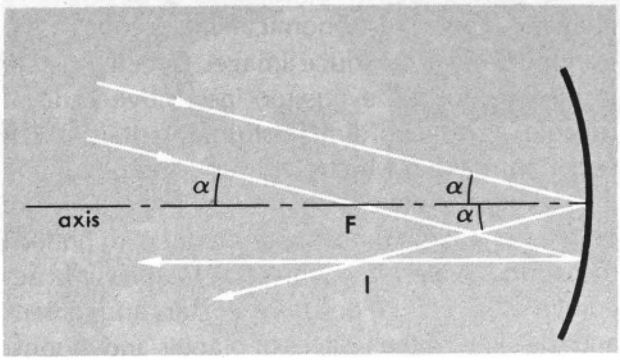
Two ways are open to answer our questions. We can do experiments with parabolic mirrors, or we can decide logically what should happen according to the two laws of reflection. Indeed, no one will stop us from using both or mixing the two methods; and we certainly must check our logical conclusions by appropriate measurements.

Let us start again with the light from a star hitting a parabolic mirror. Suppose this time that the parallel pencils of light do not come in along the axis of the mirror; instead, the direction of the rays is at an angle α to the axis. In Fig. 10 – 5 a plane is shown which includes the axis of the mirror and some of the rays representing the propagation of light from a star. By carefully constructing the reflected rays so that the angle of reflection in each case equals the angle of incidence, we can find out crudely how we expect the mirror to work. As we see in the figure, the various reflected rays pass through a small region located near the principal focus on the side of the axis opposite the star. They do not pass through a single point, because to make them do so we should have to have the axis parallel to the incident rays; but very careful construction (or calculation based on geometry) shows that the rays do come close to one point even when we include those that are not in the plane of the figure. The mirror should therefore form a blurred spot of light at the position *I* where the light almost focuses. Photographs of star images formed by a parabolic mirror do, indeed, show a slight blurring of images off the axis of the mirror.

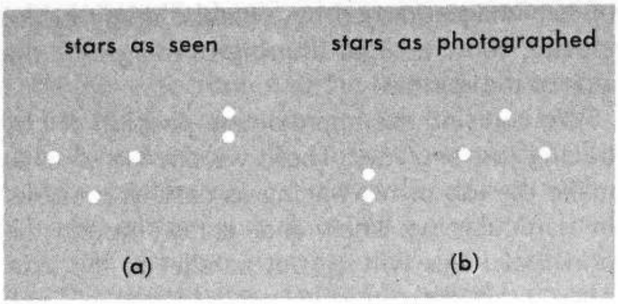
We can find the approximate position of *I* by tracing just two rays. These we choose so as to make the job of ray tracing as easy as possible. In particular we know that a ray through the principal focus will go out parallel to the axis. Also, a ray that comes in parallel to the axis will go out through the focus. And finally, a ray that hits the center of the symmetric mirror will leave it on the other side of the axis, going out on a symmetric path. Two of these rays will always allow us to find the image *I*. In Fig. 10 – 6 we locate *I* by tracing the ray that comes in through *F* (and therefore must go out parallel to the axis) and the ray that hits the mirror just at the center point. The image *I* lies at approximately the same distance from the center of the mirror as *F*, but it is at the same angle on one side of the axis that the star is on the other. Consequently, the images of several stars as seen from the center of the mirror form the same pattern as the stars themselves do in the sky;



10–5 Formation of the image of a star that is not on the axis of the telescope.



10–6 Simplified version of Fig. 10–5, showing how to find the position of the image.



10–7 A sketch of stars as seen in the sky and as photographed with a telescope. Try looking at the page upside down.

they are just turned top for bottom and right for left. A photograph taken on a photographic plate placed perpendicular to the axis at *F* shows the spacing of the stars correctly (Fig. 10 – 7). The surface on which the images are sharpest is called the *focal plane*. In reality it is not plane but very slightly curved. For the greatest precision, the photographic plate used to obtain star images is made slightly curved so that all parts of it lie in the focal plane.

10–4 THE USE OF ASTRONOMICAL TELESCOPES

A large-diameter, long-focal-length parabolic mirror used in conjunction with a photographic plate placed in its focal plane makes a very useful astronomical telescope. It gathers light from galaxies, stars, and planets and concentrates it on nearby spots in positions corresponding to the positions of these celestial objects on the celestial sphere.

A photographic film, exposed for a long time, will gather more light and record a fainter image than our eyes. It also gives a permanent record of the relative intensity and positions of star images. Large astronomical telescopes are seldom used to produce images directly visible to the eye, for the eye is too insensitive and unreliable as a recording instrument. It is much better to make a photograph and then examine the developed plate with a microscope. This not only gives a large magnification, but with built-in measuring scales it enables us to measure accurately the relative positions of stars and planets and the sizes of the images of planets and moons. Furthermore, the film can easily be scanned with a photoelectric cell to measure the relative brightness of the objects whose images are recorded on the film.

The best astronomical telescopes use very large-diameter parabolic mirrors with long focal lengths. The large diameter increases the amount of light intercepted from a distant star, forming a brighter image than can be obtained from a small-diameter mirror. Even with long exposure times, we need as much light as we can get to record images of faint stars and galaxies on a photographic plate, and so telescopes with large-diameter mirrors are used to increase the amount of light collected. The largest telescopes gather enough light so that, with very sensitive photographic plates and long exposure times, faint galaxies as far away as 10^{22} km have been recorded. The light-gathering power of large telescopes is even great enough that the light from stars can be photographed through a spectroscope attached to the telescope. Thus we observe the characteristic spectra of the elements in stars and learn that they are made of the same elements we find in our own sun and on earth.

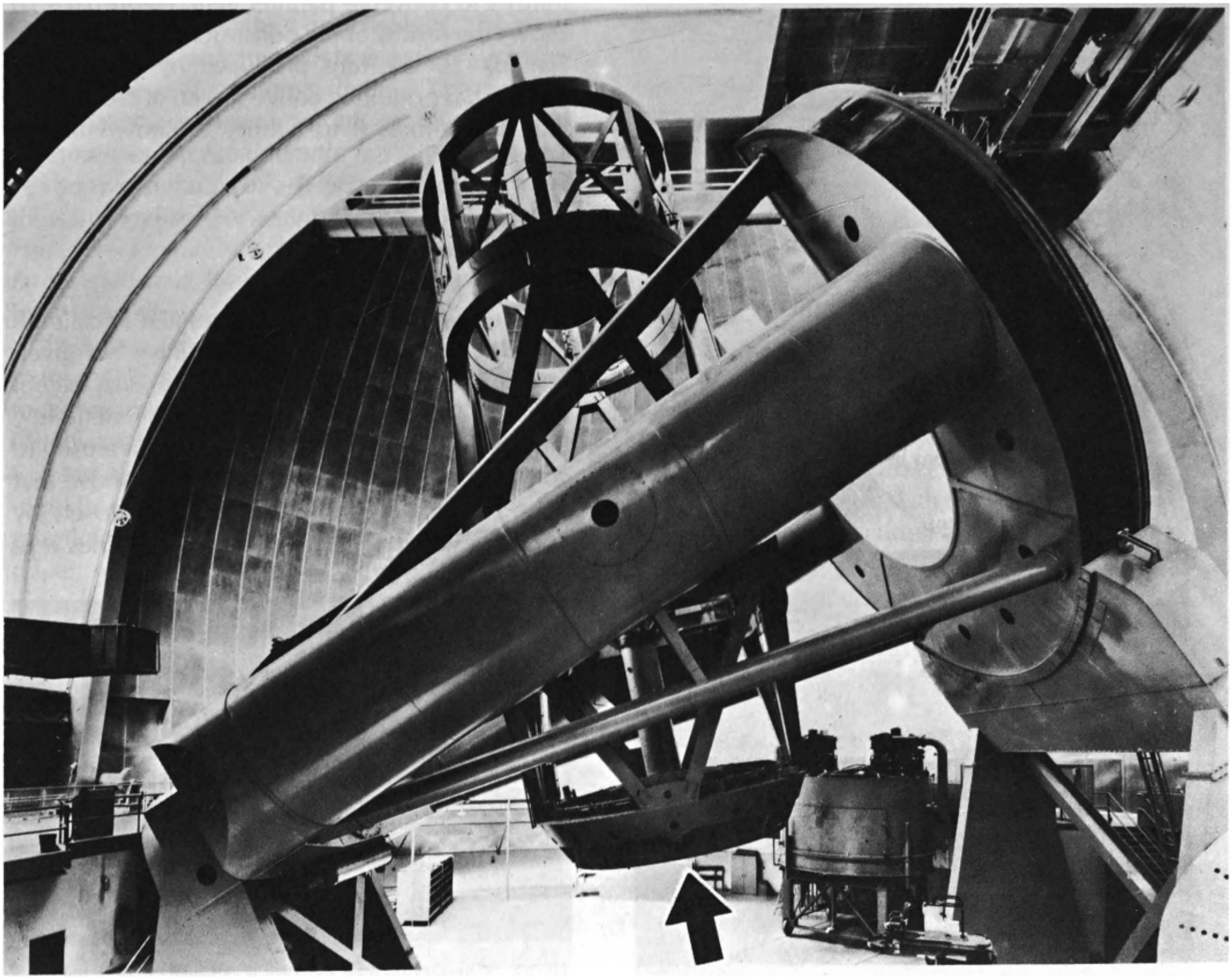
There is another advantage in using a large-diameter mirror that is just as important as its increased light-gathering power. This has to do with diffraction. You learned in Chapter 7, Section 7 – 7 that light passing through a small hole is diffracted and bent out of straight line-paths, forming a fuzzy image of the hole. In the case of an astronomical telescope, the "hole" through which the light passes is represented by the parabolic mirror. Although it is a very large "hole" indeed, it diffracts the light from distant celestial objects and makes the fine details in their images fuzzy. As a result, two stars that are very close together may never be identified as two separate stars, no matter how much magnification is used, if their images are too blurred as a result of diffraction. A large-diameter parabolic mirror gives less diffraction and better resolution of fine detail in astronomical photographs than a small one.

The longer the focal length of the mirror, the larger will be the image formed at the focus, showing more detail of the surfaces of planets and the moon, and indicating more clearly the separation of stars very close together on the celestial sphere. Of course, a long focal length and high magnification is no advantage unless good resolution is obtained by using a large-diameter mirror.

Largest in the world is the Hale telescope located on Mt. Palomar in California. The heart of the instrument is a parabolic mirror 5 meters in diameter. A metal framework which supports the parabolic mirror is mounted on axles and bearings in such a way that it can be pointed at various spots in the heavens. Figure 10 – 8 gives an idea of the mounting required. This mounting can be controlled so that it follows the apparent motion of a star from east to west while a long photographic exposure is being made.

10–5 THE CONVERGENCE OF LIGHT BY A SET OF PRISMS

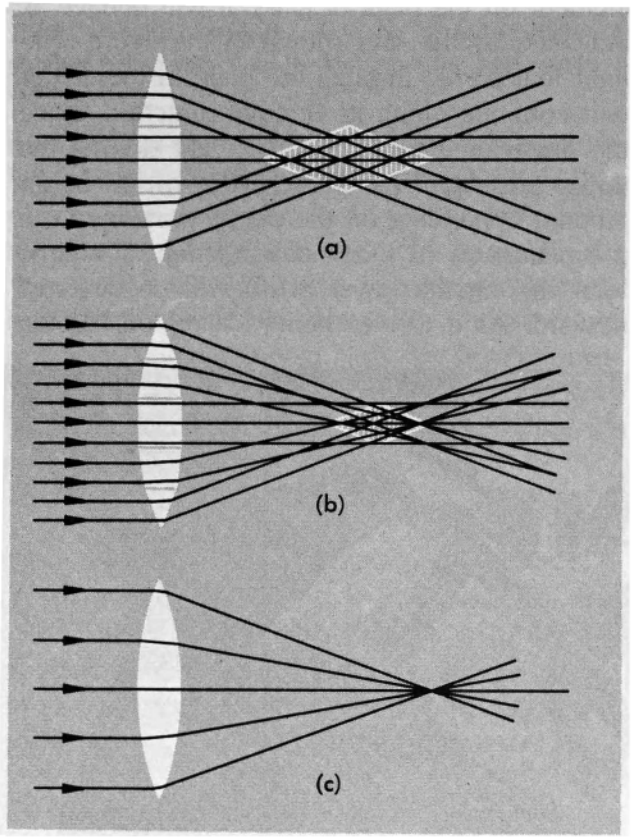
We have found that we can control and redirect light beams by the use of parabolic mirrors. Devices that can accomplish similar purposes through refraction, instead of reflection, are called lenses. To understand how a lens operates, let us examine the behavior of light in passing through the combination of a plate of glass with parallel sides and the two triangular prisms shown in Fig. 10 – 9 (a). If a parallel beam of light falls on this system from the left, so that it is normally incident on the plate of glass, it will behave as indicated by the rays shown in the figure. The light that passes through the plate in the center will continue along its original direction, since the angle of incidence is 0°. Light striking the upper prism will be deviated downward by an amount depending on the opening angle of the prism and on its index of refraction. Similarly, light striking the lower prism will be deviated upward. As a result, there is a region, shown



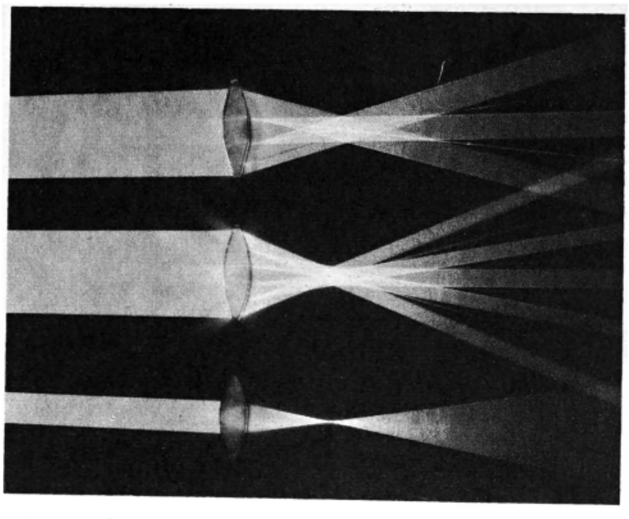
10–8 The Hale telescope at Mt. Palomar Observatory. The arrow (bottom, center) points to the mirror.

shaded in the figure, through which passes almost all of the light that falls on the plate and the prisms.

The convergence of a parallel beam of light into a limited region by this system resembles the convergence of a similar beam by a set of mirrors (see Section 10 – 1). While working with mirrors, we decreased the size of the region into which the light was converged by using an increased number of mirrors, each smaller than the original one. Let us try the same scheme here. Figure 10 – 9 (b) shows parts of the central plate and of the two prisms cut away and replaced by pieces of new prisms. The size of the shaded region is clearly smaller than it was before.



10–9 Construction of a lens by the process of subdividing prismatic sections.

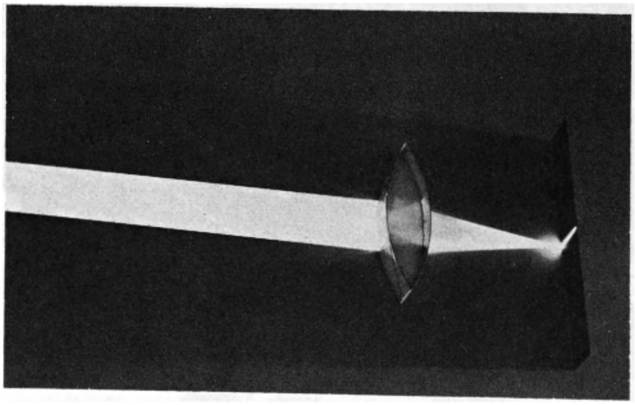


10–10 The experiments diagramed in Fig. 10–9.

If we continue the process of removing parts of the prisms and replacing them by sections having smaller opening angles, we come closer and closer to a piece of glass with the smoothly curved surface shown in Fig. 10 – 9 (c). This device is the limit that is approached as we increase the number of prisms indefinitely, just as the parabolic mirror of Fig. 10 – 3 was the limit approached as we used more and more plane mirrors to converge parallel light. Figure 10 – 10 shows the results of the construction indicated in Fig. 10 – 9. The lens produced by the process that we have outlined converges all of the parallel light that strikes it to a line, as shown in Fig. 10 – 11.

10–6 LENSES

The device we have just constructed is called a cylindrical lens. Notice that we have not given any definition of the surfaces of the lens, except that they are obtained by increasing indefinitely the number of sections of prisms that are used to converge the light. It is possible to show that these surfaces are approximated very closely by circular cylinders. In other words, the lines rep-



10–11 Convergence of light by a cylindrical lens like the one shown in Fig. 10–10. Note that the light is brought to a focus along a line.

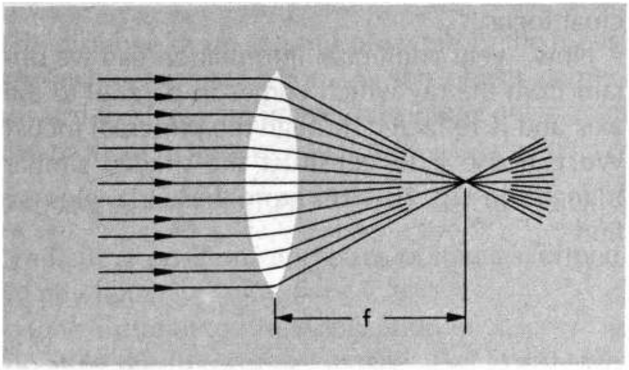
resenting the surfaces in Fig. 10 – 9 (c) are arcs of circles. The differences between the ideal surfaces and those of circular cylinders are very slight if two conditions are satisfied: (1) the thickness of the lens is considerably smaller than its width; and (2) the width of the lens is considerably smaller than the distance from the lens to the line at which parallel light is converged.

Cylindrical lenses bring the light from a distant point source of light to a focus along a line. For most purposes we prefer that the light from a point source should be focused at a point. This focusing can be accomplished by constructing a lens whose surfaces curve equally in all directions. Such surfaces are portions of spheres. Almost all lenses are bounded by two spherical surfaces.

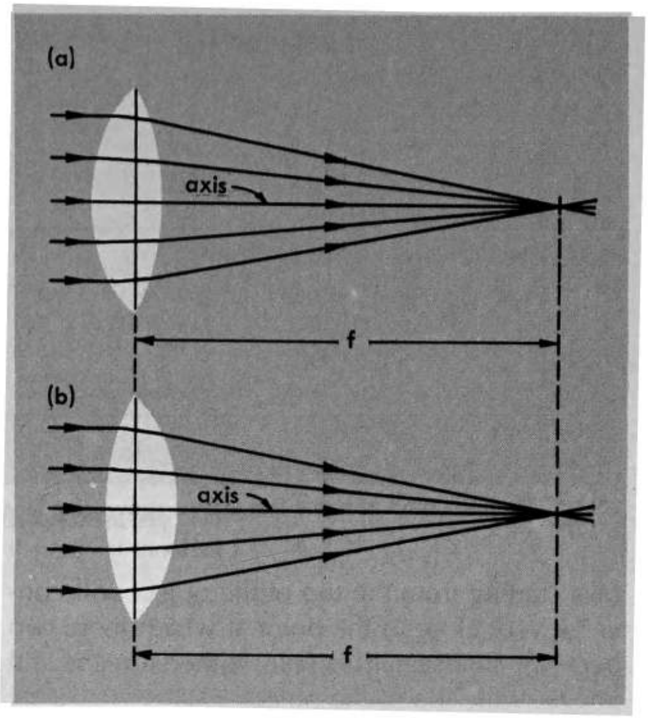
The line passing through the center of the lens and on which the centers of the two spheres are located is called the *axis* of the lens. The point on this axis at which incident parallel rays focus or converge is the *principal focus, F*. The distance of the principal focus from the center of the lens is known as the *focal length, f*.

The two surfaces of a lens do not always have the same radius. For example, the lens shown in Fig. 10 – 12 has a spherical surface of much larger radius at its right-hand boundary than it has at its left.

If a lens is thin compared to its focal length, no matter which side of the lens the light enters, the focal length is always the same. This symmetry is obvious if the lens is itself symmetric. That it is true for all thin lenses can easily be shown by an experiment in which a lens is used to focus the parallel rays of the sun to a point on a piece of paper or cardboard. If the lens is then flipped over, the focus occurs at the same distance from the lens (Fig. 10 – 13).



10–12 A lens with surfaces of unequal radii.

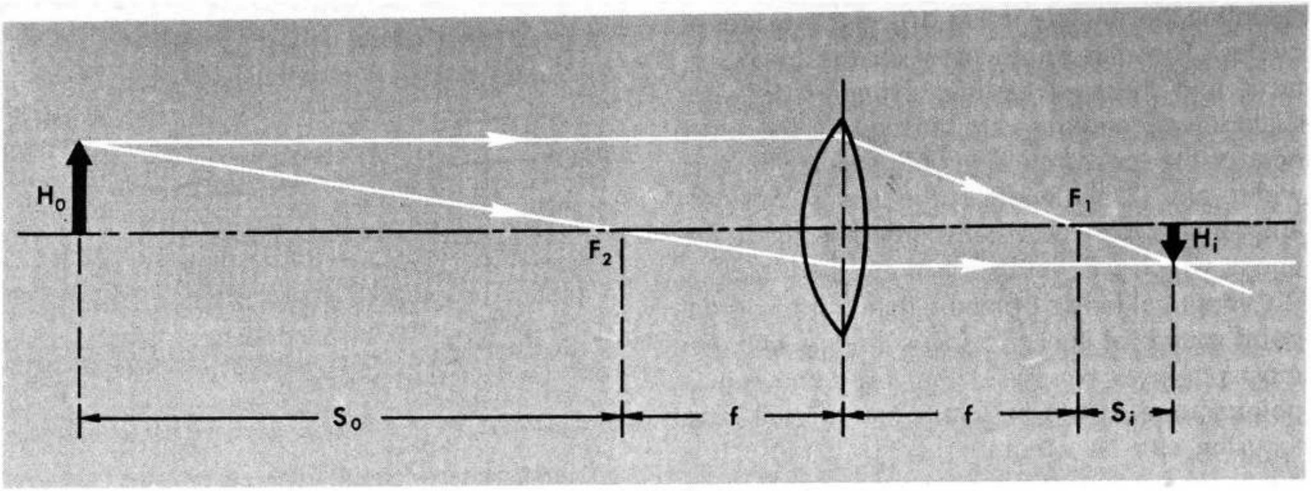


10–13 The principal foci of a lens. For thin lenses the focal distance is the same for parallel light entering either the side with a small radius (a) or the side with larger radius (b).

10–7 IMAGES FORMED BY LENSES

We have thus far concentrated our attention on the focusing of light by a lens when the light comes from a very distant object. In the practical use of lenses, we are often interested in the light coming from nearby objects, and we all know that lenses do form images of such objects. We can locate the images with the help of the knowledge that we have gained about the behavior of initially parallel rays.

Figure 10 – 14 shows a lens, an object of height H_o, and its image of height H_i. To find the location of this image, we draw the two principal rays from the top of the object, one ray parallel to the axis of the lens and the other through the principal focus F_2. The ray parallel to the axis is bent by the lens so as to pass through the principal focus F_1. We also know that rays coming from the right and the parallel to the axis would be deviated to pass through the other principal focus F_2. It follows from the reversibility of light paths that the ray from the top of the object that passes through F_2 from the left must travel parallel to the axis after it has passed through the lens. All



10–14 The formation of a real image by a converging lens.

rays starting from the top of the object will converge very close to the point at which these two bent rays intersect. This point is therefore the real image of the top of the object.

We see an image of the top of the object because the light moving along various paths is sent through an image point and diverges from there to our eye or camera just as if a small version of the object were located there. The intersection of the two rays tells not only how far away from F_1 the image of the top is, but also determines the length H_i of the whole image as it extends down from the axis. By drawing such a two-ray picture to scale, we can always find out what the laws of refraction predict about the position and size of the image.

We could have chosen any other point on the object and located its image in the same way. Had we done so for a number of points, we would have found that the image falls along the line that is shown in the figure.

You probably noticed that, in constructing the two principal rays, we did not consider the exact path of the ray within the lens, but broke it sharply. This approximate construction is good enough for our present purposes because our location of the two principal foci is accurate only if the lens thickness (at its center) is small compared with the focal length. The only lenses to which our construction accurately applies are therefore *thin lenses*. For the purposes of ray diagrams, we may consider such lenses to be circular plates perpendicular to the axis.

The geometry of these two-ray drawings is so simple that it hardly seems necessary to go further in specifying the relations. But it is a nuisance to draw diagrams all the time, and the simplicity of the geometry suggests that we can express our results in extremely simple forms for numerical computation. We shall start with the relation of image to object size.

In Fig. 10 – 15 we have left out everything but the ray from the top of the object going through the principal focus. Because this ray is refracted parallel to the axis and passes through the image, the "top" of the image must lie the same distance away from the axis as the refracted portion of this ray. Therefore, H_i has the same height as the perpendicular from F to the refracted ray. Now we can get the ratio of H_i and H_o immediately from the shaded similar triangles. Because f and S_o are the bases and H_i and H_o the corresponding altitudes, we get

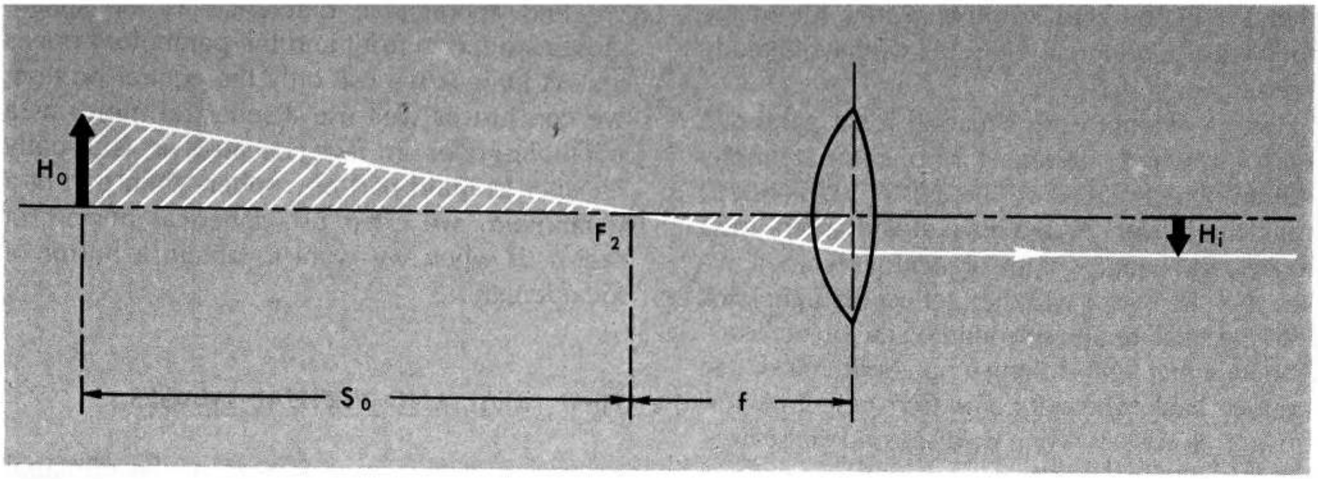
$$\frac{H_i}{H_o} = \frac{f}{S_o}$$

To determine the ratio of image size to object size, all we need to know is the ratio of focal length to the distance of the object from the principal focus, F_1.

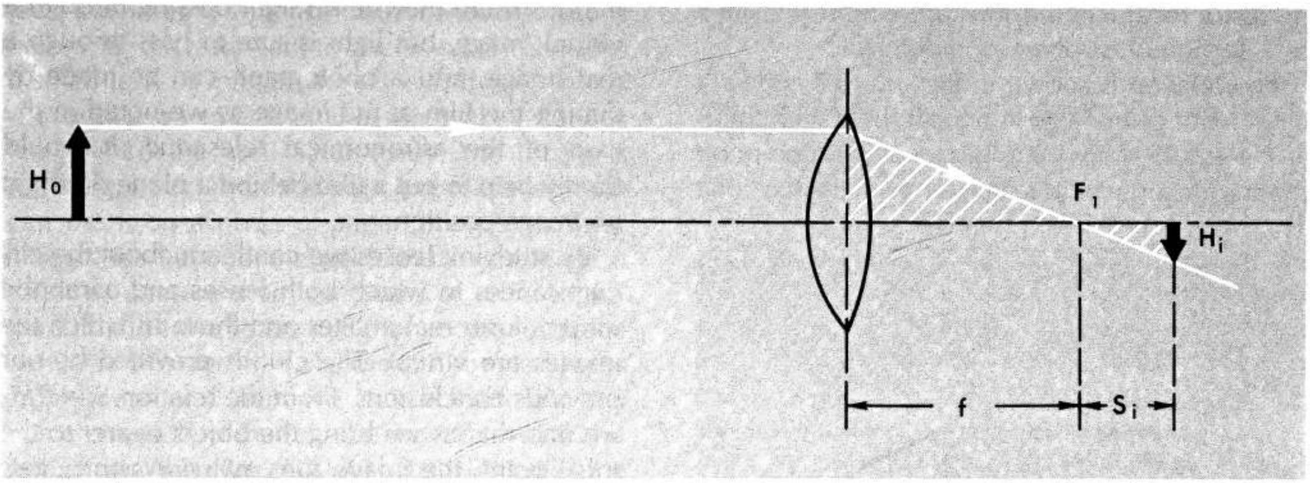
Now, what additional information can we obtain from the ray which comes in parallel to the axis and is refracted through the principal focus? We can use it to construct the shaded similar triangles in Fig. 10 – 16. From these triangles we find

$$\frac{H_i}{H_o} = \frac{S_i}{f}$$

When this relation is combined with $\frac{H_i}{H_o} = \frac{f}{S_o}$, we



10–15 Finding the ratio of image size to object size by similar triangles. The ray from the top of the object through the principal focus is used to relate the ratio of the sizes to the ratio of the focal length and object distance.



10–16 Finding another relation between image and object size. The ray from the top of the object parallel to the axis is used to relate the ratio of the sizes to the ratio of focal length and image distance.

get $\frac{S_i}{f} = \frac{f}{S_o}$

$$\text{or } S_iS_o = f^2$$

The distance of image and object from the focus are in inverse proportion. As the object comes toward the lens, the image must move out so that the product S_oS_i stays equal to f^2.

Using $S_i = \frac{f^2}{S_o}$ and $H_i = \frac{f}{S_o}H_o$, we can quickly predict the positions and sizes of images formed by any lens.

One example of the application of $S_oS_i = f^2$ is too attractive to resist. Let us ask: at what distance from the principal focus must the object be placed so that the image will be the same distance from the lens as the object? That is, when are S_o and S_i equal? Clearly the answer is: when both S_o and S_i are equal to f; that is, when object and image are both located at a distance $2f$ from the lens. Furthermore, image and object are the same size, for

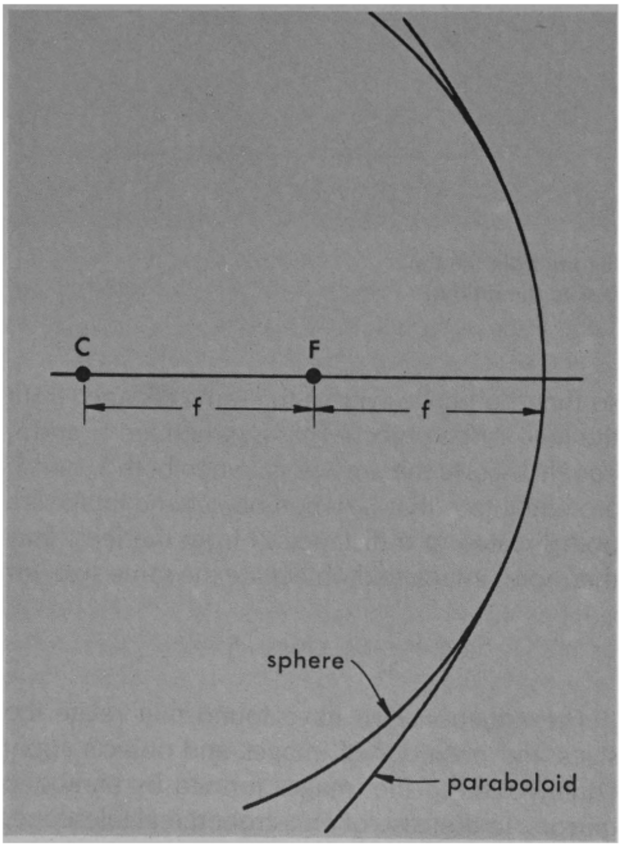
$$H_i = \frac{f}{S_o}H_o \text{ gives } H_i = H_o$$

The equations we have found that relate the sizes and positions of images and objects apply equally well to the images formed by parabolic mirrors. In the case of an astronomical telescope, for example, where S_o is, for all practical purposes, infinite, we see by the equation $S_iS_o = f^2$

that S_i must be zero. In other words, the image forms at the principal focus, as we have already seen in Section 10 – 3.

Let us examine the situation for a parabolic mirror in which the object is not a star at infinity but an object placed on the axis at a distance $2f$ from the mirror. As we have already seen, for a lens S_o will equal S_i. If the object is very small, we see that the mirror will then reflect the light back into the small region surrounding the object itself. The idea that light striking a parabolic mirror can bounce back where it came from might be surprising. As we saw when we discussed the searchlight, a spherical mirror will bounce light coming from the point at the center of the sphere right back to that point. Apparently, the effective portion of a parabolic mirror must closely resemble a portion of a sphere. In particular, a parabolic mirror of focal length f must be almost the same as a section of a sphere of radius $2f$.

The relation is shown in Fig. 10 – 17, where a circle with center at C and a parabola with focus at F are both drawn through the central point on the mirror surface. It is only far from the axis that we find appreciable difference as the sphere closes around in front and the paraboloid opens up. As long as we use only the central portion, we can just as well use a spherical mirror as a parabolic one. In fact, because it is usually cheaper to build part of a sphere than part of a paraboloid, we often use spherical mirrors of radius $2f$ when we want to obtain a mirror of focal length f.



10–17 The relation of spherical and parabolic mirrors. The common principal focus is at F, and the center of the sphere is at C.

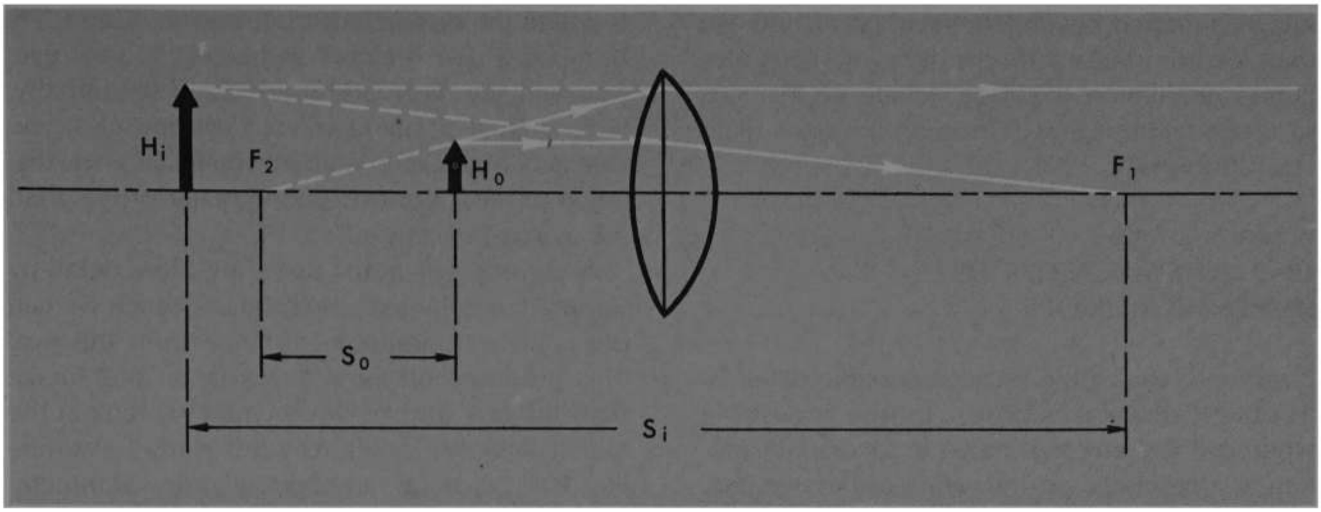
10–8 REAL AND VIRTUAL IMAGES

An image of the moon formed at the principal focus of the parabolic mirror in a telescope is a *real image*. The light really converges to it and diverges from it to our eyes. It is thus distinguished from the virtual image seen in a plane mirror. There may be no light passing through a virtual image, but light is sure to pass through a real image, and a photograph can be made by putting the film at the image as we noted in the case of the astronomical telescope. It would hardly help to put a film behind a plane mirror at an image position.

By studying lenses we can learn about the circumstances in which both lenses and parabolic mirrors form real images and those in which the images are virtual. The clue is provided by our previous conclusions. From the relation $S_i = f^2/S_o$ we find that as we bring the object nearer to the focal point, the image runs away. When S_o gets very small, S_i, the image distance, becomes huge; and as the object passes through the principal focus, the real image disappears at an infinite distance from the lens.

If we continue to move the object closer to the lens, we are not surprised to find a virtual image approaching from an infinite distance on the other side of the lens and reaching the lens when the object does.

A two-ray diagram (Fig. 10 – 18) now makes it easy to see that whenever the object is between the principal focus and the lens we get a virtual image. We use the same two rays — the *principal rays,* they are called — the one parallel to the axis, which must refract through the focus, and the one coming from the focus to the lens and then going out parallel to the axis. As Fig. 10 – 18 shows, after refraction these rays appear to come from an intersection behind the object, and the image is always magnified compared to the ob-



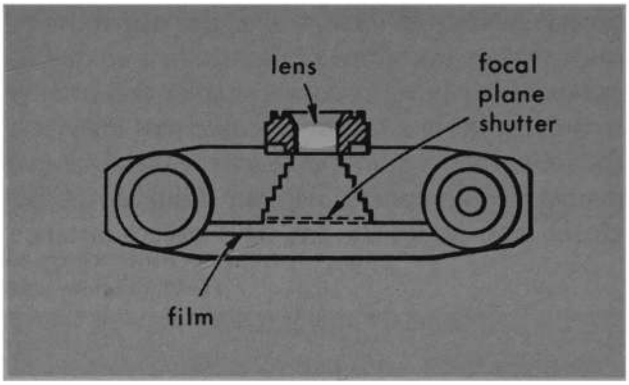
10–18 Formation of a virtual image in a convex lens. The object lies between the principal focus F and the lens. The principal rays are used to find the expected location of the image.

ject. Ray tracing will work out the details just as before; and if you want to check experimentally, a small lens or magnifying glass will do.

Produce an image of the sun with a convex lens. Since the sun is far away, the image is formed practically at the principal focus and you can see it on a piece of paper. Images of closer objects lie beyond the principal focus, and in order to capture them on paper or on a photographic film we have to change the distance between lens and film. To make a simple photographic camera, then, we need a light-tight box containing a lens that can be moved toward and away from the film (Fig. 10 – 19). By adjusting the position of the lens, we can place a sharp image on the photographic film.

A projector is just a camera worked backward. You can make one by taking the back off a camera, mounting a slide where the film usually goes, and shining a bright light through the slide and out through the lens. The lens then forms an enlarged image well in front of the camera, where you can place a screen.

In cameras, projectors, and other man-made optical instruments, images are always brought into focus by changing the position of a lens with respect to the object. The eye, on the other hand, works differently: it focuses images on the retina by changing its curvature and hence the focal length of its lens. When an object is at a very large distance from the eye, the rays entering the eye are nearly parallel and an image is formed at the principal focus as shown in Fig. 10 – 20 (a). When a nearby object is viewed, the



10–19 A camera. The distance between the lens and the film is adjusted to give a sharp image on the film.

10–20 The lens of an eye adjusted to focus the light from a distant object (a) and from one near by.

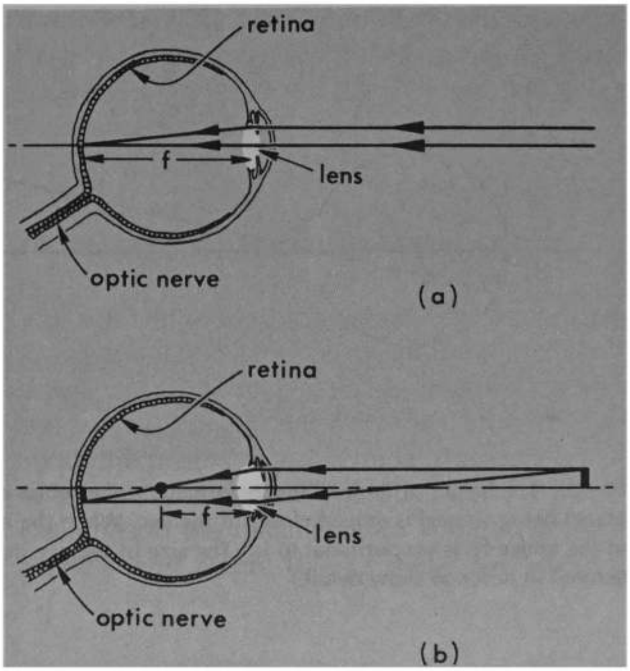
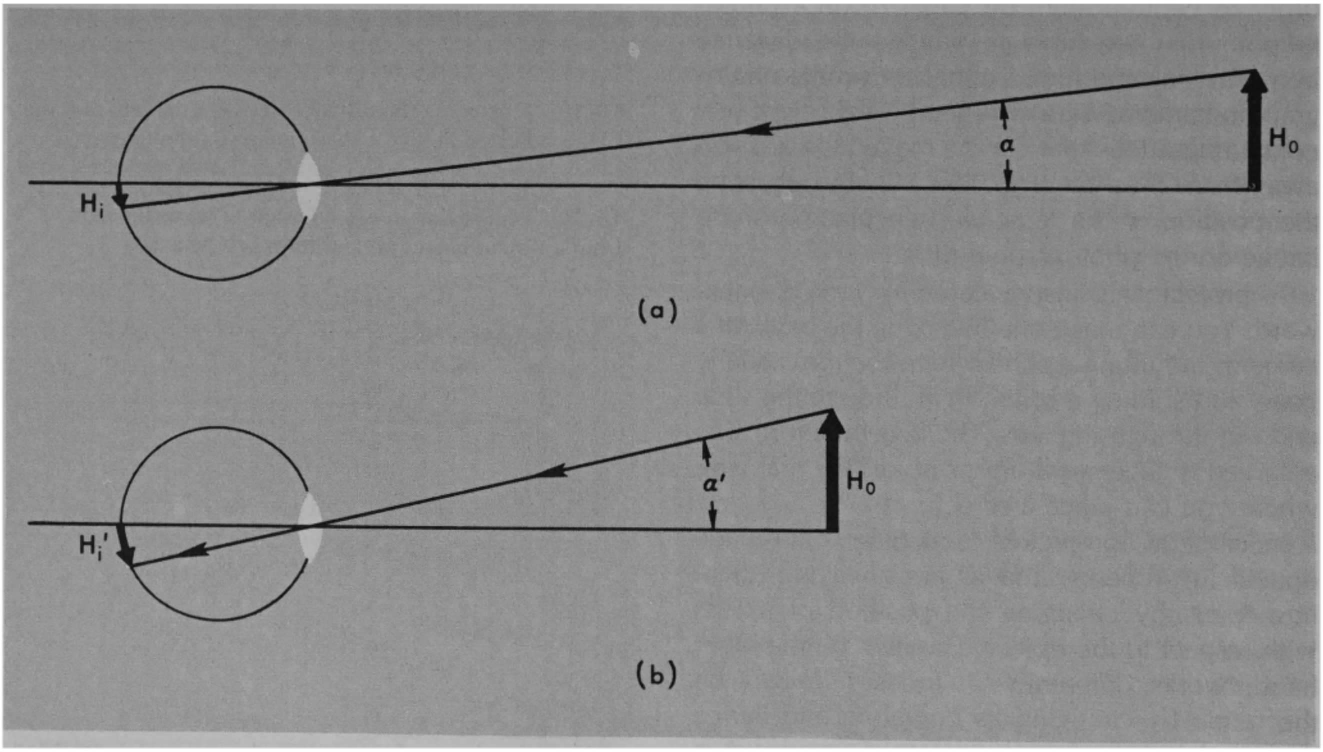


image is formed beyond the focal point, and eye muscles from the elastic eye lens into a sharper curve, decreasing the focal length of the lens so that a real image will form on the retina (Fig. 10 – 20 (b).

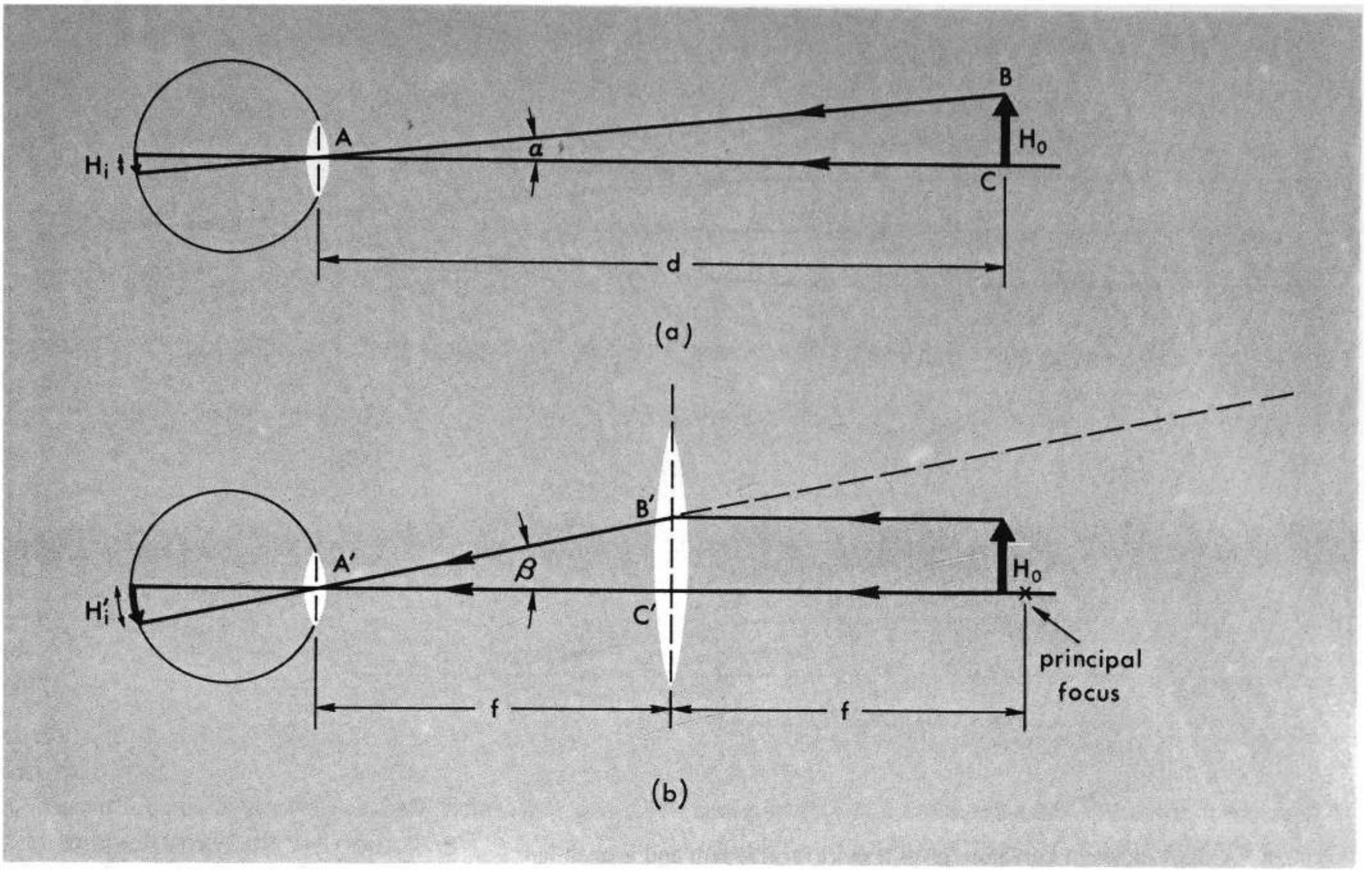
10–9 THE MAGNIFIER OR SIMPLE MICROSCOPE

If we wish to see the greatest possible detail of an object, we get it as close to the eye as possible while still keeping the image in focus, thus getting a large image on the retina of the eye (Fig. 10 – 21). The size of this image is proportional to the angle α so long as α is small, as is usually the case. But there is a limit to how close to the eye an object can be and still be viewed. As the object gets closer to the eye, the eye muscles must change the shape of the eye lens so that its radius of curvature becomes smaller and smaller in order to form a sharply focused real image on the retina. Soon a limit is reached; the adult eye cannot accommodate, without strain, an object closer than about 25 cm. This object distance is called the *distance of most distinct vision*. Try bringing a pencil closer and closer to your eye. You will see more and more detail until finally, with a great straining of your eye muscles, you can no longer keep a sharp image. Is your distance of most distinct vision greater or less than the average of 25 cm?

A convex lens helps us to see more detail by forming an enlarged virtual image which we can place at a comfortable distance from the eye. This produces on the retina a larger real image than the one formed there when we look at the object directly. To see why this is true, examine Fig. 10 – 22. In (a), an object H_o viewed directly at the closest comfortable distance d forms an image whose size is H_i on the retina. If a lens of focal length f is introduced between the eye and the object, and the object is just inside the principal focus of the lens as shown in Fig. 10 – 22 (b), a larger image H_i forms on the retina. H_i' is larger than H_i by the ratio $\frac{H_i'}{H_i} \approx \frac{\beta}{\alpha}$ and this ratio we can define as the magnification produced by the lens when used in this manner as a magnifier or simple microscope.



10–21 The height of the real image formed on the retina of the eye increases as an object being viewed is moved closer to the eye. When the angle α is small, the height of the image H_i is proportional to it. (The size of the eye in this figure has been exaggerated in order to show detail.)



10–22 A converging lens placed between a nearby object and the eye increases the size of the real image formed on the retina of the eye. (In this figure, as in Fig. 10–21, the size of the eye has been exaggerated in order to show detail.)

The ratio $\frac{\beta}{\alpha}$ in Fig. 10 – 22 depends only on the focal length *f* of the lens, as you can easily see by comparing the sides *AC* and *A′C′* in the triangles *ABC* and *A′B′C′* in the figure. If α and β are small,

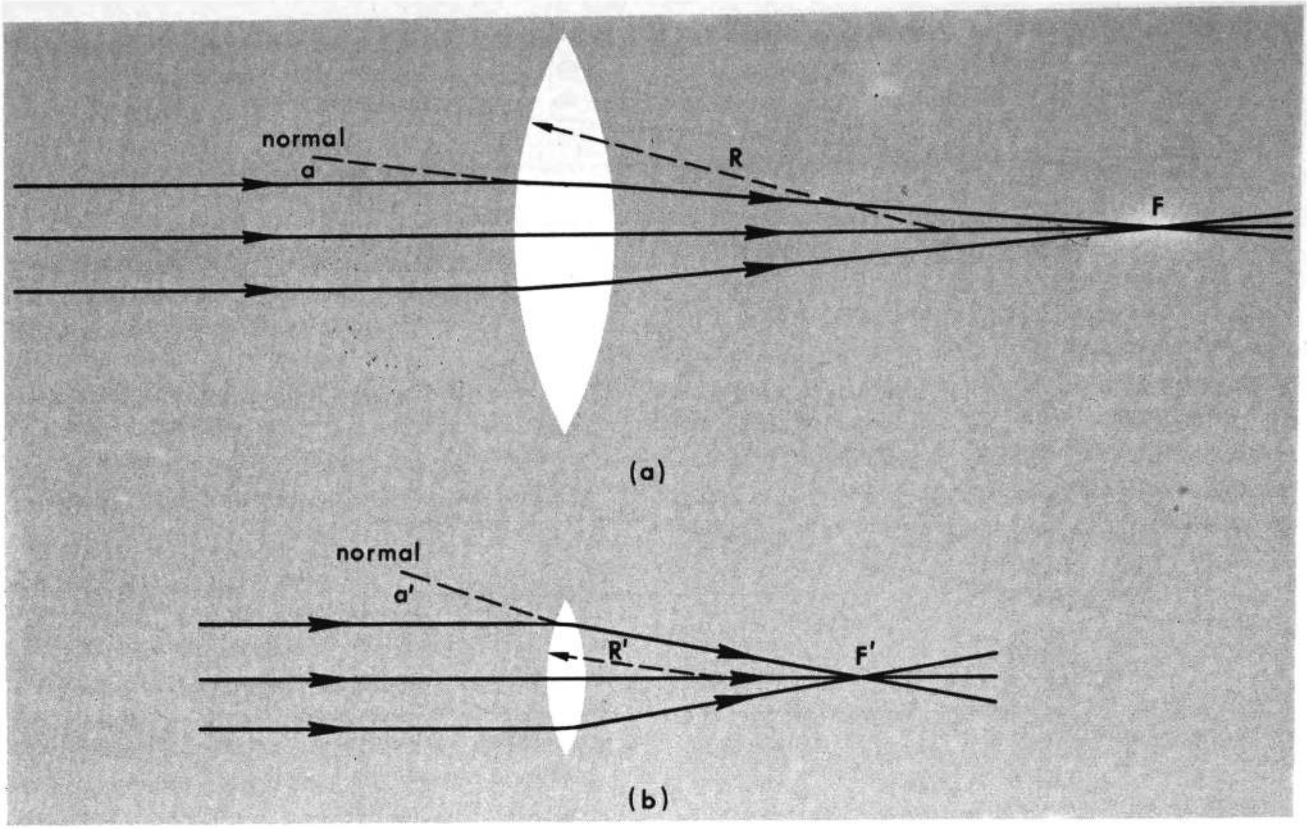
$$\text{magnification} = \frac{\beta}{\alpha} = \frac{A'C'}{AC} = \frac{d}{f}$$

and *d*, the distance of distinct vision, is always about 25 cm.

Notice that no matter where the object is placed (so long as it is between the principal focus and the lens to give a virtual image) the angle β remains the same and the magnification does not change. Of course, if we place the object so close to the lens that the virtual image formed by the lens is closer to the eye than about 25 cm, the eye lens cannot adjust itself enough to give a sharp real image on the retina, so it is best to avoid eyestrain by having the object close to the principal focus with the virtual image a large distance to the right of the lens. Under these circumstances, the eye does not have to be at the focus of the lens; the magnification will stay the same so long as the virtual image is very far away compared to the distance from the eye to the lens. In practice, in order to obtain a large magnification, a lens with a focal length much shorter than $d = 25$ cm is used, and the lens is held close to the eye.

10–10 THE COMPOUND MICROSCOPE

You may think that it is possible to get as large a magnification as you wish, without limit, simply by using a magnifying lens with a very short focal length. However, a little thought about the factors that determine the focal length of a lens will convince you that there are practical limits to the magnification you can get with a single lens. Figure 10 – 23 shows two lenses of



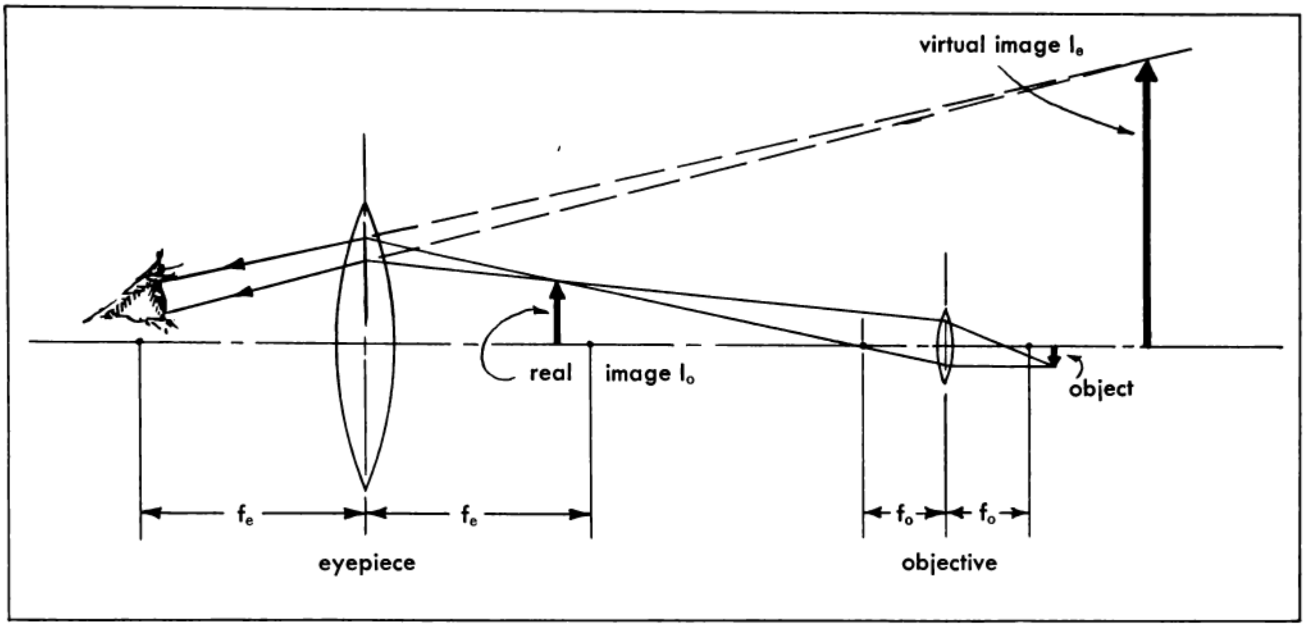
10–23 A short radius of curvature gives a short focal length and a small lens.

different radii of curvature R and R'. The α' at which the light is incident on the lens in (b) is larger than the angle α in (a), so, by Snell's law, a light ray is bent more entering the lens in (b) than the corresponding ray that strikes the face of the lens in (a). As we see, the focal length of the lens in (b) which has the shorter radius of curvature is less than the focal length of the lens in (a) which has a larger radius of curvature. Of course, the light leaving the lenses is bent again, increasing the effect of the smaller radius in (b).

As you can see, the size of the lens in (b) is smaller than that in (a), and we have increased the magnification at the expense of the amount of light we can get through the lens and the size of the field of view we can see through the lens. Even if we try to increase the size of a small lens by using a whole sphere instead of two small slices of a sphere, back-to-back, we encounter other troubles. (See Section 10 – 12.)

For a given size of lens, we can decrease the focal length a little by making the lens out of a material with a high index of refraction so that the bending will be greater for a given angle of incidence on the lens surface. But this will not help much since diamond, which has the highest index of refraction of transparent materials, has an index only 4/3 that of glass.

A way out of this dilemma is first to use a big lens to form an enlarged real image and then to view this already enlarged image with a magnifier as shown in Fig. 10 – 24. The total magnification will then be the product of the magnification of the objective (the lens that produces the enlarged real image) and the eyepiece (the lens used as a magnifier to give an enlarged virtual image of the real image). Such a combination of lenses is called a compound microscope and is the most common and useful type of optical microscope today. Perhaps you have used one in a biology course. It consists of a long tube fitted with a converging lens at each end. The bottom lens, called the *objective lens,* is of short focal length, and the microscope is adjusted until the object is just beyond the principal focus of the objective lens so that it forms an enlarged real image, I_0. This image is at such a distance from the eyepiece (slightly nearer the eyepiece than its principal focus) that a



10–24 **Compound microscope. The eyepiece and objective are located at the two ends of the tube. The diagram illustrates the formation of the images.**

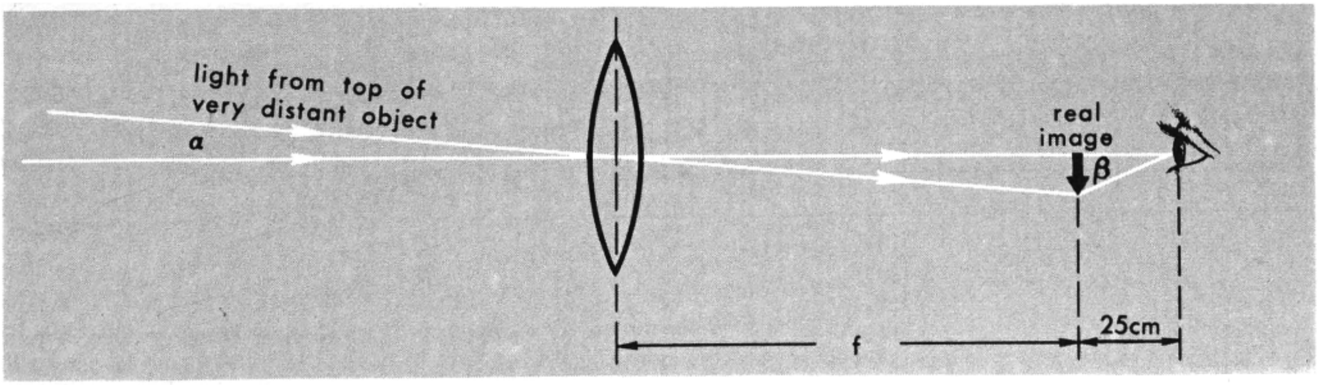
virtual image, I_e, is formed at the distance of most distinct vision. The eyepiece is thus used as a simple microscope to examine the enlarged real image formed by the objective.

10–11 REFRACTING TELESCOPES

As we have seen in Section 10 – 9, when we move closer to a distant object it forms a larger image in the eye and we can see more detail. When we cannot move closer to a distant object, such as a mountain top, we can produce a larger image on the retina of the eye by using a converging lens. Lenses used in this way to form larger eye images of distant objects are called *refracting telescopes.*

The simplest kind of telescope consists of a single lens as shown in Fig. 10 – 25 where a distant object forms a real image at the principal focus of the lens. As you can easily see, α is the angle the light from the top of the distant object would make when entering the eye directly, and the magnification β/α is equal to $\frac{f}{25}$ where f is the focal length of the lens in centimeters. Even though the real image at the focus of the lens is much smaller than the distant object, in terms of what the eye sees, the image on the retina has been increased in size by β/α over what the eye sees viewing the distant object directly.

As with the simple microscope or magnifying glass, there are practical limits to the magnification that can be obtained with this simple telescope. To obtain a magnification of 100 times, for example, would require a lens with



10–25 **The formation of an image by a single-lens telescope.**

a focal length of 25 cm × 100 = 2500 cm = 25 m, and a telescope 25 m long is large and unmanageable, to say the least. But we can use the same trick we used to improve the simple microscope. We need only to add an eyepiece and use it to magnify the real image. Then the overall magnification becomes the product of the magnification of the long-focal-length objective lens and the magnification of the eyepiece used as a simple microscope; that is,

$$M = \frac{F}{25\text{ cm}} \times \frac{25\text{ cm}}{f} = \frac{F}{f}$$

We now have a telescope which will give a large magnification with an objective of reasonable focal length if we use an eyepiece whose focal length is much less than that of the objective.

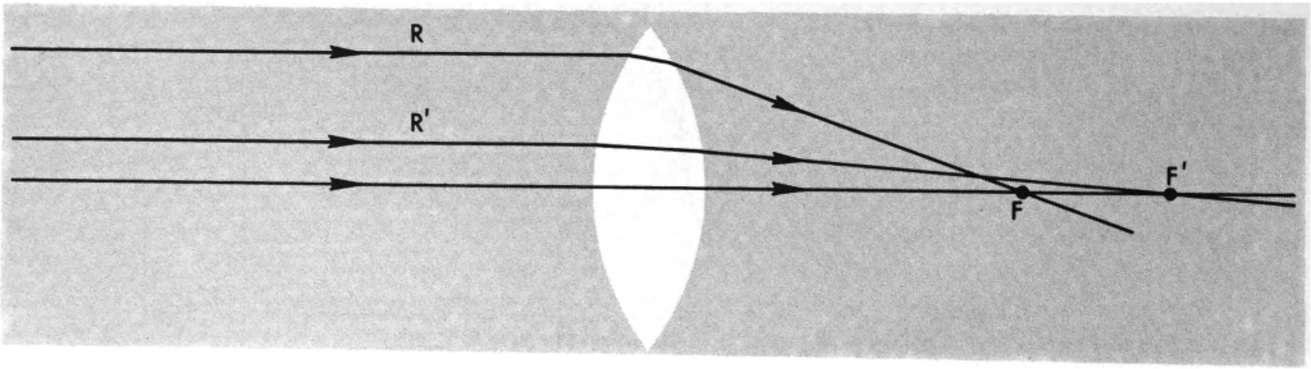
Of course, if we use an objective lens of long focal length, its surfaces will have a large radius of curvature, and we can make the diameter of the objective very large. This is of particular advantage in an astronomical telescope where we wish to collect as much light as possible in order to see very distant, and therefore dim, stars and galaxies. Furthermore, a single-lens telescope, with a film at the principal focus, can be used as a sort of giant camera and be fixed on a star for a long time. Thus, light from stars much too dim to be seen will expose the films. The developed film can then be examined with a microscope to give an increased magnification.

10–12 LIMITATIONS OF OPTICAL INSTRUMENTS: ABERRATIONS AND RESOLUTIONS

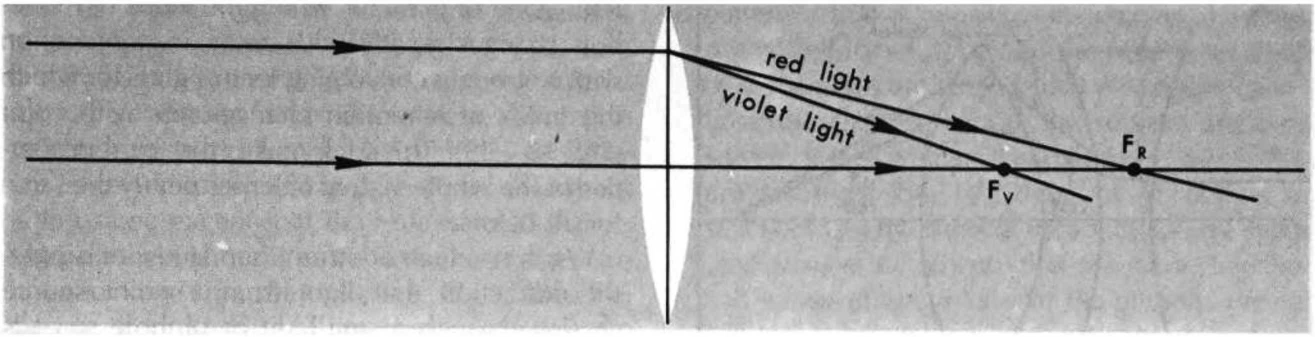
The simple camera, simple and compound microscopes and telescopes as we have described them all have limitations that reduce their usefulness. We have used the physical laws of reflection and refraction to design instruments to perform specific, useful tasks, aiding our natural, inborn senses. But we cannot get photographs giving as sharp detail as we should like, nor as large magnifications from microscopes and telescopes.

We have already seen (Fig. 10 – 17) that in order to get clear images from either spherical or parabolic mirrors, we must not use too much of the sphere or parabola. The same restriction applies to lenses whose surfaces are spherical. (Parabolic surfaces are never used in lenses, since they are not so good as spherical surfaces, even for distant objects located along the axis.) One way to see why using too much of a sphere leads to fuzzy images is to follow the paths of light rays that locate the principal focus, both for rays close to the axis and rays far from the axis as shown in Fig. 10 – 26. The figure shows that the position of the principal focus, for both spherical mirrors and spherical lenses, is different for pencils of light close to the axis and for those far away from the axis where the surface closes around the axis appreciably. Thus, images formed by different parts of the surface fall at slightly different positions with different magnifications. This effect of spherical surfaces on the reflection and refraction of light is called *spherical aberration* and leads to fuzzy and distorted images. As you can see, light near the axis comes to a focus at more nearly a common point than the light far from the axis. In other words, using only a small part of the spherical surfaces gives clearer, sharper images.

In the case of lenses, there is another kind of aberration due to the fact that the index of re-



10–26 The refraction of light by a lens having spherical aberration.



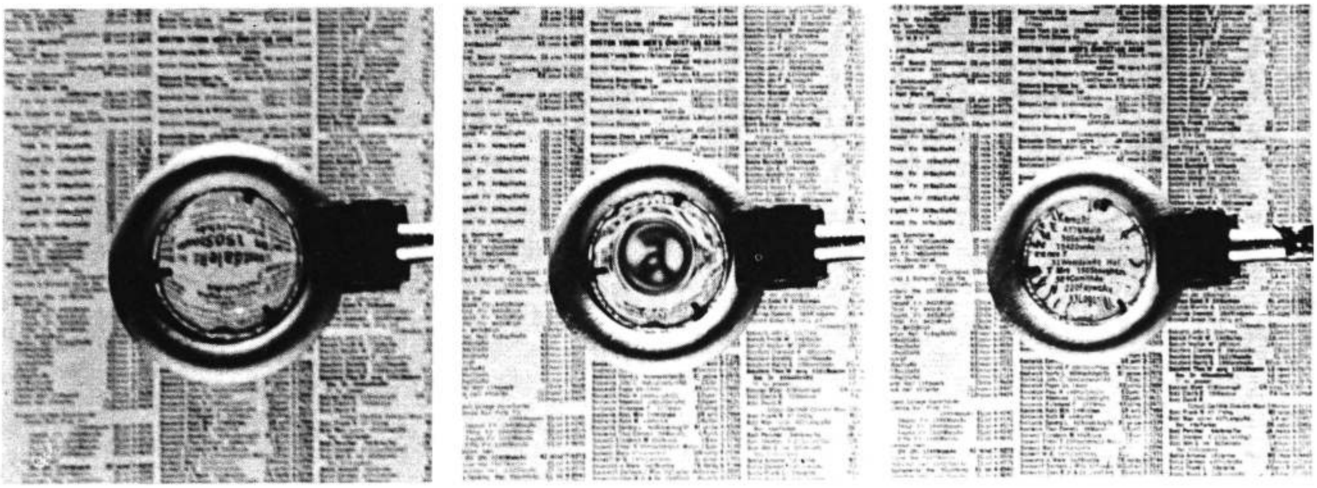
10–27 The refraction of white light by a lens having chromatic aberration.

fraction is different for different wavelengths of light. Figure 10 – 27 shows how dispersion in glass gives rise to different principal foci for light of different colors, resulting in different positions for images formed by light of different colors. As with spherical aberration, chromatic aberration can be reduced by using only a small, central portion of a lens. In photography, distortion and blurring are often reduced by using a "stop"—a barrier with a small hole in it—so that only the central portion of the lens is used. But this works only in bright light. In dim light the whole lens must be used to get enough light to expose the film, and under these conditions the resulting photograph is not so sharp.

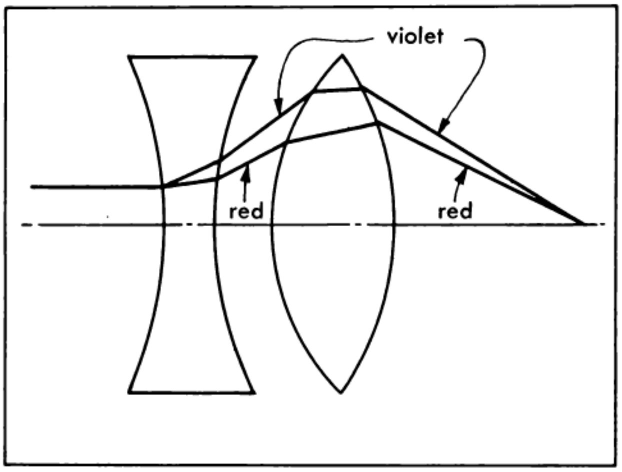
To see the effects of aberrations on images, you can do the following experiment:

If you hold a magnifying glass close to this page, you can see a clear, undistorted, and slightly magnified image of the print. Now slowly raise the glass from the paper and at the same time increase the distance between your eyes and the lens. At some positions the image appears distorted. If your eyes are far enough from the glass, you may also detect some rainbow colors when looking at a corner of the page. In Fig. 10 – 28 we see three pictures. They were made by looking through the same lens. In the first picture the lens was held so that the page of the telephone book lies entirely (but slightly) below the focal region; in the second picture the page is in the focal region; and in the last the focal region lies below the page. Clearly, in each case the image of the page looks quite different from the page itself. Part of the game of designing really good optical instruments is to minimize the geometrical distortions so apparent in these pictures.

Both spherical and chromatic aberration can be greatly reduced by using a weakly diverging



10–28 Distortion by a lens. These three photographs were made by looking through the same lens. At left, the lens was held so that the page of the telephone book is slightly below the focal region; in the middle picture, the page is in the focal region; at right, the focal region lies below the page. Note the geometrical distortions.

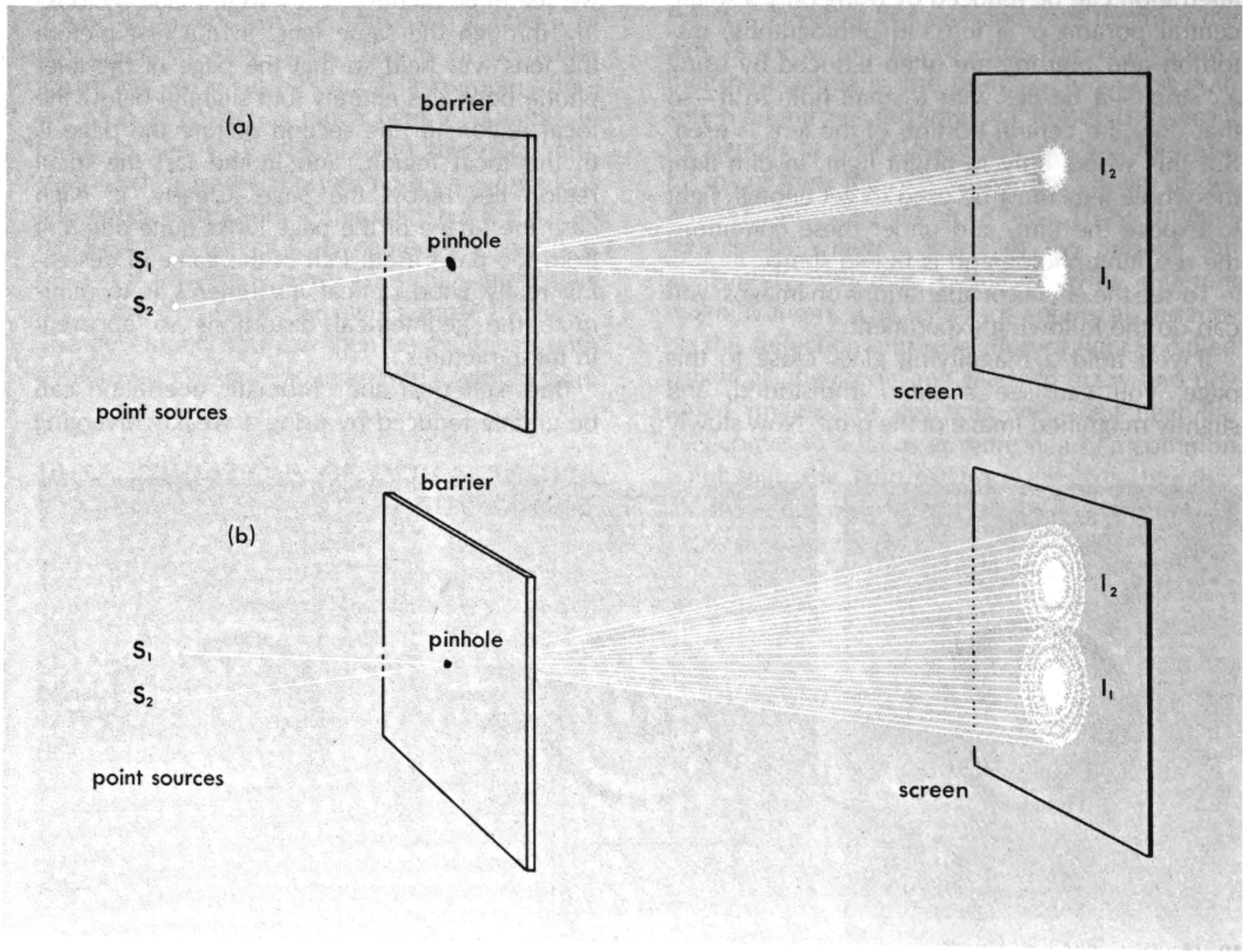


10–29 A lens built of two pieces to minimize the different focal properties of light of different colors. Such doublets are often made with one common surface and glued together. They are called achromatic lenses.

lens made of glass for which the index of refraction changes greatly with color, in conjunction with a strongly converging lens of glass for which the index of refraction changes less with color (Fig. 10 – 29). This trick makes the focal properties of the whole system of lenses nearly the same for all colors.

We have learned from our study of single-slit diffraction that light from a point source passing through a small slit or pinhole spreads out to give an image larger than the size of the hole. We also learned that small holes spread the light more than large holes. This diffraction caused by a small hole is of great importance in designing microscopes and telescopes. It determines the ultimate limit of their magnification.

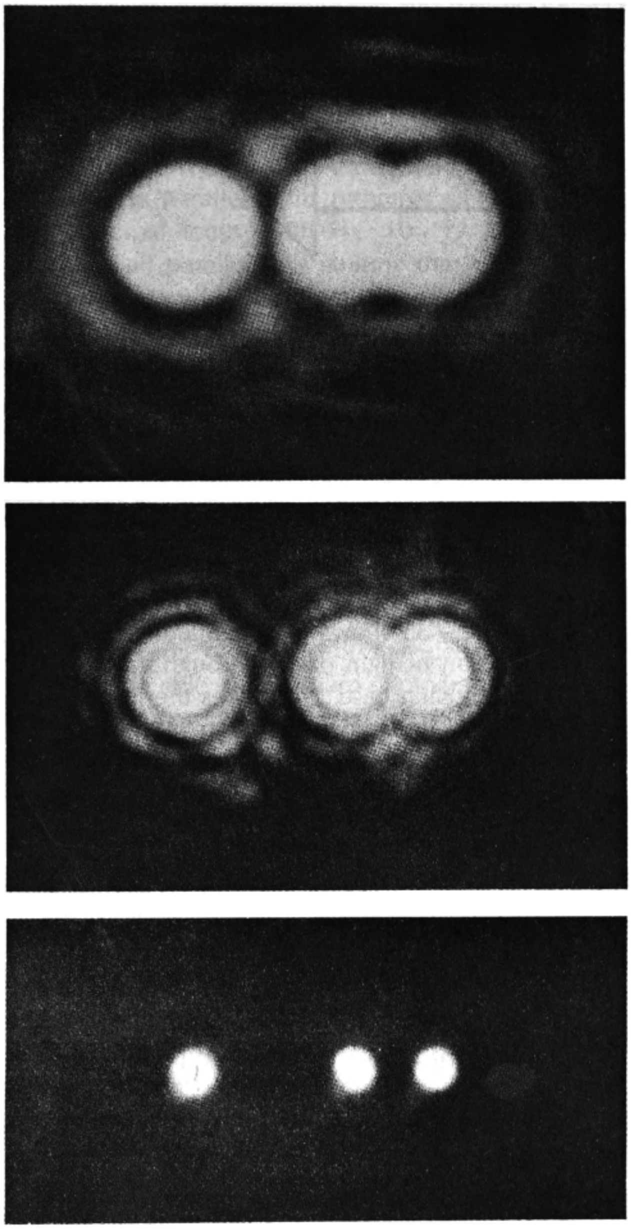
To see why diffraction limits magnification, let us consider what happens when two point



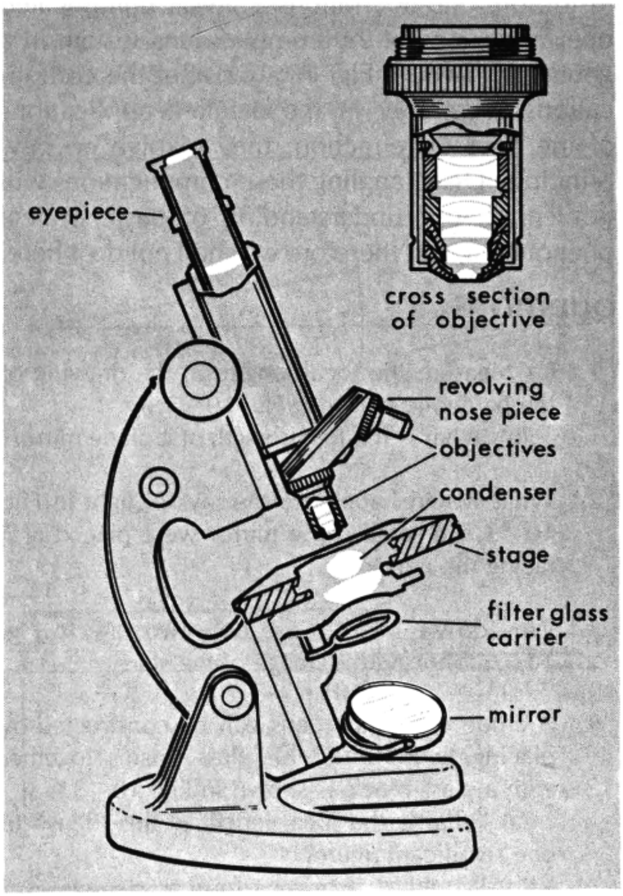
10–30 (a) When light from the sources S_1 and S_2 passes through a small hole, the images are fuzzy but resolved. (b) When the hole is very small, the images overlap and are unresolved (see Fig. 10–31).

sources, close together, send light through a pinhole onto a screen. In Fig. 10 – 30 (a), I_1 is the image of source S_1, and I_2 is the image of source S_2. These images are, in fact, diffraction patterns produced by the tiny pinhole, and they are large and fuzzy. In Fig. 10 – 30 (b) everything is the same except that the hole is smaller. This spreads the light still more, and the images are now so large that they overlap. It is difficult to decide, from looking at the screen, whether the pattern is that of two separate sources or a single odd-shaped source. When the hole is so small, or the sources are so close that the images cannot be distinguished, we say the sources are *unresolved.* When we can separate them, we say they are *resolved.* The *resolution* of an optical instrument is a measure of its ability to give separated images of objects that are close together.

If we substitute a lens for the pinhole, we can focus the light from two point sources and produce what appear to be sharp images. Careful examination of these images, however, shows that a lens cannot eliminate the spreading of light by diffraction, because light in passing through a lens is passing through a hole of limited size. Figure 10 – 31 shows a series of photographs of three point sources seen through lenses of the same focal length but different diameter. In the top picture the light passes through a small-diameter lens; in the second picture it passes through a large lens; and in the last, through a very large lens. In the top picture, the images are unresolved. If a larger lens is



10–31 Diffraction patterns from three point sources produced by lenses of different diameters. As the diameter is increased (top to bottom), the resolution improves. (Courtesy: Francis Sears, "Principles of Physics, III," Addison-Wesley Publishing Co.)



10–32 A modern compound microscope.

used, diffraction decreases and resolution improves, until with a very large lens the images are so clearly separated that even if the sources were closer together we could still resolve them.

In a telescope, the greater the size of the objective, whether mirror or lens, the better the resolution. The large Hale telescope on Mt. Palomar gives sufficiently good resolution to resolve individual stars in the Andromeda Nebula, a feat that smaller telescopes are unable to accomplish.

In a microscope, where the light passes through a small objective lens, we expect diffraction to be important. We can usefully increase the magnification of a microscope until diffraction prevents us from resolving nearby objects. A further increase in magnification makes everything look larger, but it does not enable us to resolve objects which are close together. Increasing the magnifying power of microscopes with the best of lenses is futile beyond the point where detail fails to be resolved because of diffraction.

The problem of designing a system of lenses with the best resolution and smallest amount of distortions and aberrations is a very complicated one, as shown by the complex lens system in a good microscope (Fig. 10 – 32). But the complications arise only in the detailed applications of the laws of refraction; they involve no new principle. Disentangling these complications will not enrich our understanding of basic optical phenomena, and therefore we shall not do it here.

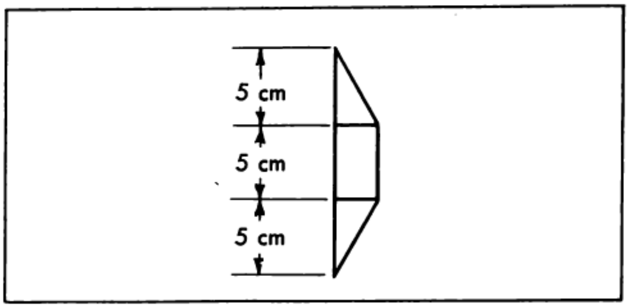
QUESTIONS

1.* (a) Measure the focal length on the drawing of the parabolic mirror in Fig. 10 – 3.
(b) What is the focal length of a plane mirror?

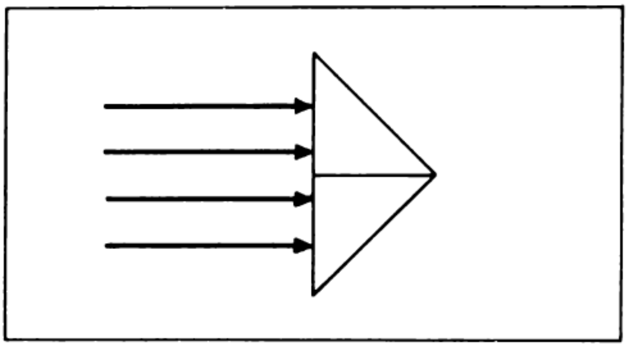
2.* What would happen to the rays of light in Fig. 10 – 3 if a small plane mirror were placed at F facing the mirror?

3.* Why do we need to use only two rays in Fig. 10 – 6 to locate the image of the star?

4. A crude converging lens can be constructed by placing two 30°-60°-90° glass prisms together with a glass block as shown in Fig. 10 – 33.
(a) What is the focal length of this "lens" to one significant figure?
(b) Would such a lens form a clear image? Explain.



10–33 For Question 4.



10-34 For Question 5.

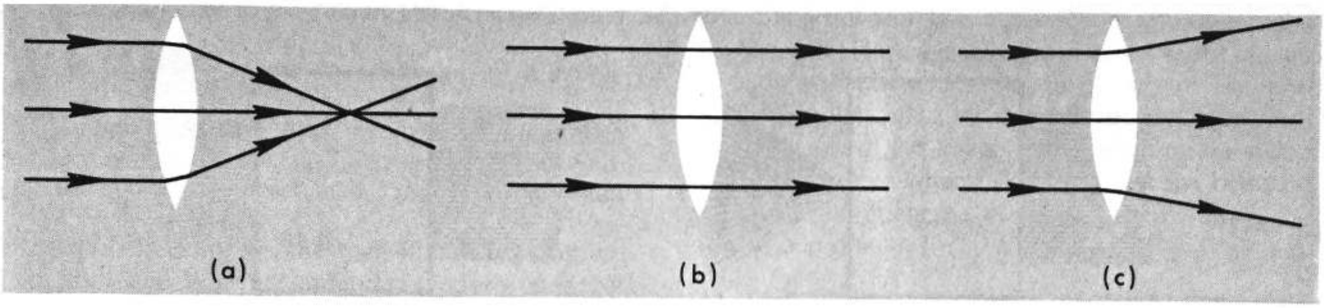
5. If two 45° prisms of glass (index = 1.50) are arranged as in Fig. 10 – 34 they will not converge parallel light. Why not? What will happen to the light?

6.* A lens is placed in three liquids one after another. Figure 10 – 35 shows what happens when a parallel beam of light passes through the lens and liquid in each case. How does the index of refraction of each liquid compare with that of the lens?

7. A nail 4.0 cm high stands in front of a converging lens at a distance of 15 cm from the principal focus. The focal length of the lens is 10 cm. What is the size of the image?

8. The distances of an object and its image in a converging lens are often measured from the center of the lens, instead of from the principal focus. We call these distances D_o and D_i respectively. We have $S_o = D_o - f$ and $S_i = D_i - f$ where f is the focal length. Using these relations, show that from $S_oS_i = f^2$ follows $\frac{1}{D_o} + \frac{1}{D_i} = \frac{1}{f}$

9. The image of a candle is 30 cm from the center of a converging lens. The candle is 10 cm long and its image is 5.0 cm long. What is the focal length of the lens?



10–35 For Question 6.

10. How large an image of the sun will be formed by the Palomar telescope, whose focal length is 18 m? The sun's diameter is about 1.4×10^9 m, and it is 1.5×10^{11} m away.

11. Compare Fig. 10 – 3, a paraboloid which reflects all rays parallel to its principal axis through its principal focus, with Fig. 10 – 17 and explain why all parallel rays striking the surface of the sphere would not reflect through F.

12. Be prepared to discuss what happens to the image as the object is moved along the principal axis of a converging lens from infinity to the lens.

13.* Is a real image formed by a parabolic mirror ever larger than the actual object?

14. A camera lens of focal length 5 cm is mounted on a fine screw thread for adjusting the focus. What range of movement is needed if the camera is to be able to take sharp photographs at 1 m as well as from great distances?

15. A single thin lens of 30 cm focal length is placed in front of the lens of a camera; the camera lens is set at "infinity." Where should you place an object to get a clear picture on the film with the extra lens?

16. If a camera is focused to give a sharp image of a point object at a certain distance, the images of point objects at greater or lesser distances will be small circular patches.

(a) For points on the axis of the lens, draw ray diagrams to show this effect.

(b) If the circular patch becomes larger than about 0.25 mm diameter, the image of the particular point will begin to appear blurred ("out of focus"). If a camera has a lens of 10 cm focal length and 3 cm diameter and is focused sharply for objects 5.1 m away, what are the distances of the nearest and farthest objects that will be in tolerable focus at the same setting? (Check your result, if possible, against a camera that shows "depth of focus" markings.)

17. A lens whose focal length is 10 cm is used in a slide projector to give a real image on a screen at a distance of 6.0 meters.

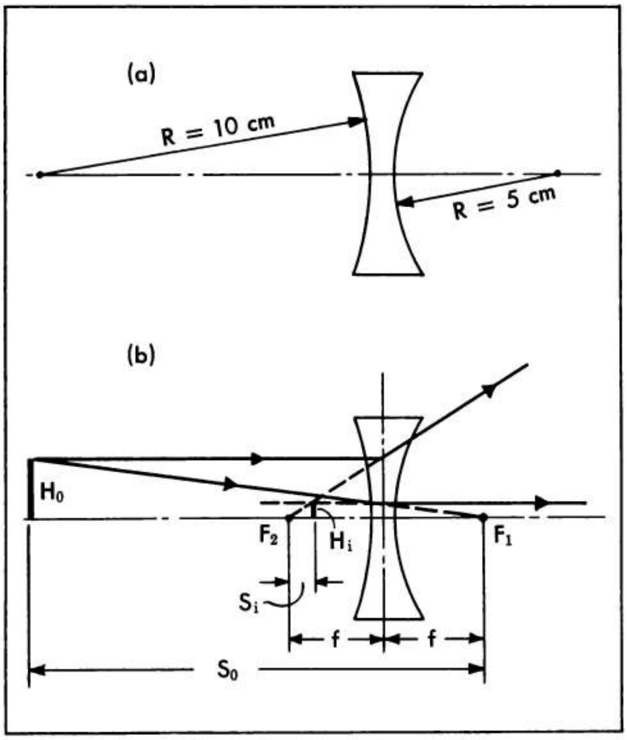
(a) What will be the magnification?

(b) How far is the lens placed from the slide?

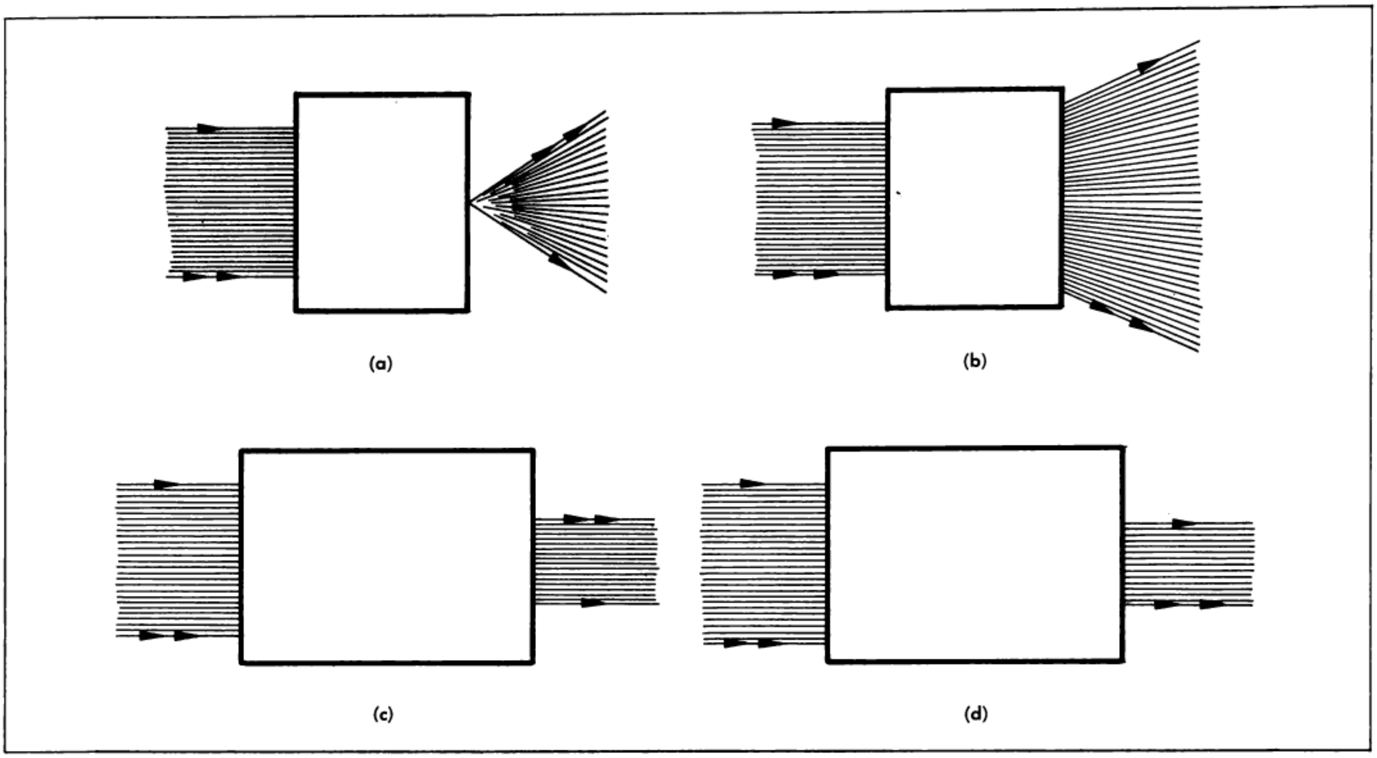
18. (a) From the ray diagram of Fig. 10 – 36 show that $S_0 S_i = f^2$. Notice from which focal points S_0 and S_i are measured.

(b) What happens as you move the object toward the lens? Can S_i ever get bigger than S_0? Is the image ever bigger than the object?

(c) How would you find (experimentally) the focal length of a diverging lens?



10–36 For Question 18.



10–37 For Question 19.

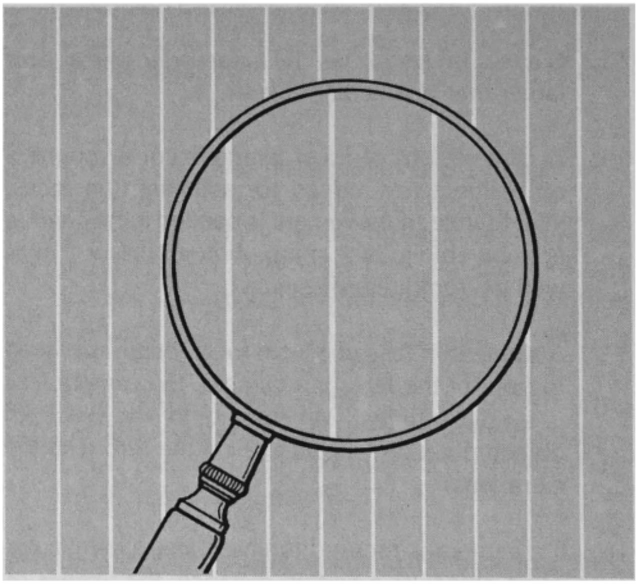
19. In Fig. 10 – 37, a parallel beam of monochromatic light enters each box from the left. Draw what could be in each box to produce the effects shown. The single and double arrows on the emerging beam show the corresponding edges of the entering beam.

20. (a) What is the angle of vision of a tree 12 m high at a distance of 100 m?
 (b) Does the angle of vision of the tree depend on how high above the ground the observer's eye is?

21. Suppose someone walks toward you from a distance of 60 m to a distance of 20 m. What is the ratio of the angle of vision of his head at 20 m to the angle of vision at 60 m?

22. A certain person cannot clearly and comfortably see objects placed closer to his eyes than 80 cm. If he wishes to read a book held 25 cm in front of his eyes, what focal distance does he need for his glasses? The lenses are 1.5 cm in front of the eyes.

23. To measure the focal length of a thin lens a person held it 4.0 cm in front of a lined paper and kept his eye 25 cm above the paper looking through the lens. He could then also see the paper directly. (See Fig. 10 – 38.)
 What focal length did he find?



10–38 For Question 23.

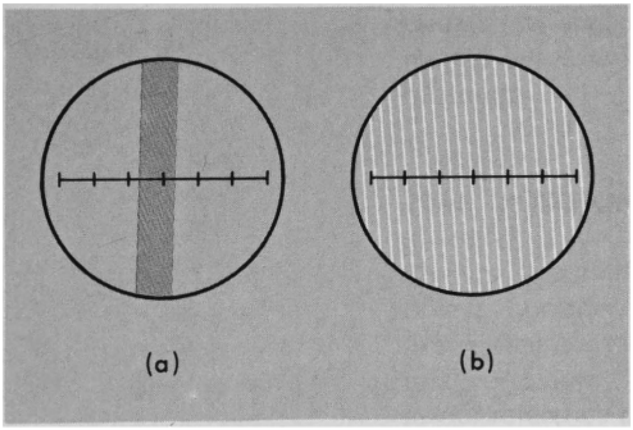
24. For the magnification of an eyepiece, we found $\frac{d}{f}$ where d is taken as the distance of most distinct (or closest distinct) vision. If your eyes can

accommodate to see directly at 15 cm, we should write $\frac{15\text{ cm}}{f}$ as the magnification of a simple magnifier for you. Also, if you can accommodate to images no closer than 35 cm, $\frac{35\text{ cm}}{f}$ would apply. Why does the magnification go up for someone who accommodates poorly at small distances? Does he see more detail than someone who can accommodate closer? Be prepared to discuss this question in class.

25. Assume that your distance of most distinct vision is 25 cm. A compound microscope has an eyepiece of 2.0 cm focal length and an objective of 4.0 mm focal length. The distance between objective and eyepiece is 22.3 cm. What is its magnification?

26. Using the microscope of Question 25, with the same adjustment, we see an amoeba. With a ruler, we measure the size of the virtual image by looking at it with one eye and at the ruler with the other. On the ruler the amoeba appears to be about 10 cm long. About how big is it really?

27. Suppose you wish to measure the diameter of a very thin thread. To do this you look at the thread through a microscope in which there is a scale. The scale is placed so that you can clearly observe the scale and the thread at the same time (see Fig. 10 – 39(a)).
   (a) Where must the scale be placed?
   You then look through the microscope at a diffraction grating (Fig. 10 – 39(b)). When used in a spectroscope the grating diffracts monochromatic light at an angle of 26° when the wavelength is 5900Å.
   (b) How thick is the thread?



10–39 For Question 27.

28. Suppose you use a telescope to observe the moon. The objective (the lens nearest to the object) has a focal length of 30 cm. The eyepiece lens has a focal length of 5 cm. How does the size of the image of the moon on the retina of your eye when using the telescope compare with the image on the retina when viewing the moon directly?

29. A photo is taken of a distant object using a camera having a lens with an extra long focal length of 25 cm. After developing the film, a lens whose focal length is 2.0 cm is held close to the eye in examining the photograph. How many times larger is the angle of vision of the image viewed through the lens compared to the angle of vision when the distant object was observed directly?

30. Sun spots can easily be observed if an image of the sun 10 cm in diameter is projected on a screen. Where would you place the screen and lenses *A* and *B*, of 30 and 1.0 cm focal length, respectively, to produce such an image? (The sun subtends an angle of 0.54°, or about 0.0094 radians.)

31. You have three lenses, two of focal length 5 cm and one of focal length 25 cm.
   (a) How would you combine these lenses to make a telescope in which the landscape is seen right side up?
   (b) Why do you need only two lenses to make an astronomical telescope?

32. The temperature of an oven is shown on a thermometer that can be read clearly when 0.5 m from the eyes. However, owing to the intense heat it has to be read from a distance of 4 m, using a telescope.
   (a) What magnification is needed?
   (b) Will a telescope marked "magnification 7 times" do?
   (Magnification 7 times means that the telescope is focused at infinity; that is, both object and image are at infinity.)

33.* How does the magnification in the top picture in Fig. 10 – 31 compare with that in the bottom picture?

34. How large an image will be formed of an artificial satellite 53 cm in diameter passing at an altitude of 500 miles, if it is photographed with a camera whose lens has a focal length of 10 cm? Would you expect an actual photograph to show a larger or smaller image than the size you have calculated?

35. Two images can just be resolved when the central maximum of one falls on the first node of the other.

    (a) Show that the resolution of a narrow slit depends upon λ/w where w is the slit width.

    (b) About how close together can two line sources be placed and still be resolved if they are viewed through a .01-cm slit (about the smallest size you can make easily) 3 meters away from the sources?

36. Microscopes in which the object is illuminated by ultraviolet light can give higher magnifications than microscopes that use visible light.

    (a) How do you explain this?

    (b) How are the images seen if no visible light is used?

    (c) Since glass is opaque to ultraviolet light, how can such a microscope be made?

37. Stars are often photographed through a blue filter. What is the advantage of this?

11

Motion Along a Straight-line Path

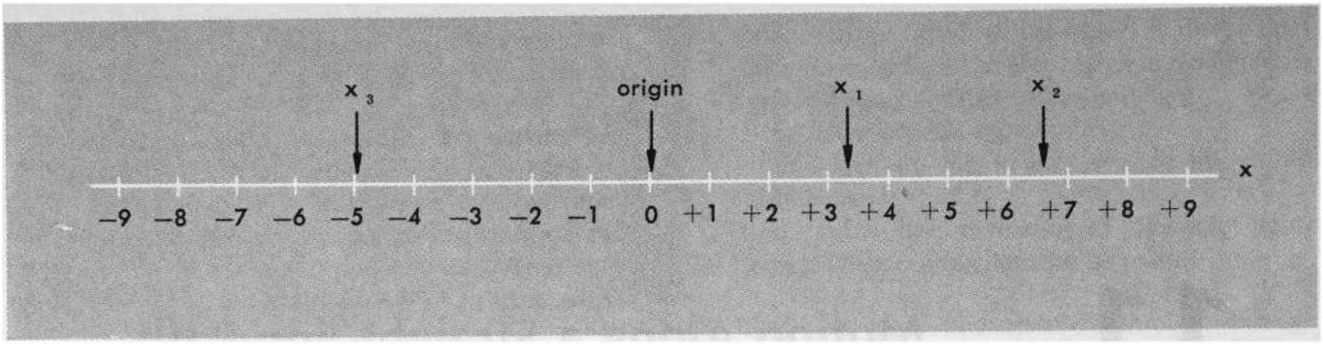
A FREIGHT TRAIN is rolling down the track at 60 kilometers per hour. Out of the fog a mile behind, a fast express appears, going at 110 kilometers per hour on the same track. The express engineer slams on his brakes. With the brakes set he needs three kilometers to stop. Will there be a crash? What we are called upon to do here is to predict where the two trains will be at subsequent times, and to find in particular whether they are ever at the same place at the same time. In a more general sense, we are asking about the connections between speeds, positions, and times.

The general subject of such relationships is called *kinematics*. In studying kinematics we do not concern ourselves with questions such as "Why does the express train need three kilometers to stop?" To answer such a question we would need to study in detail how the brakes slow down the train. Such questions as these will be considered in later chapters. In this chapter and the next we shall consider only the description of motion. We shall start with the discussion of motion along a straight-line path. Then we shall extend the discussion to describe more general motions.

In both of these chapters we shall draw on our ability to measure time and position, for all motion is the changing of position as time goes on. Usually, we shall not think consciously of the time and position measurements, but without them we would in fact be talking words without meaning.

11–1 POSITION AND DISPLACEMENT ALONG A LINE

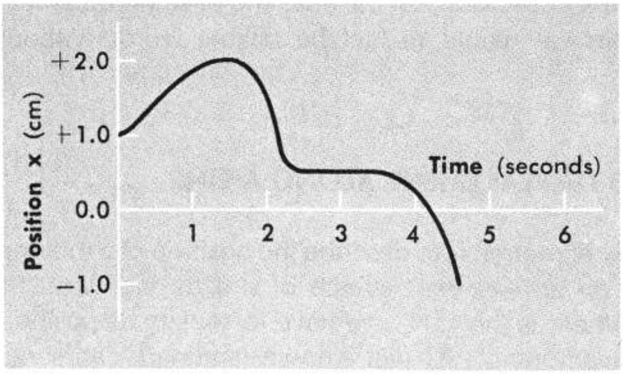
The first step in the study of motion is to describe the position of a moving object. Consider a car on an east-west stretch of straight highway. To answer the question "Where is the car?" we have to specify its position relative to some particular point. Any well-known landmark can serve as our reference point, or origin, for measuring position. The choice of a landmark is arbitrary so long as it is clearly stated and understood by



11–1 The x coordinate line.

everyone. We then state how far the car is from the landmark and in which direction, east or west, and the description of position is complete. Thus, for example, we say that the car is 5 kilometers west of the center of town, or it is 3 kilometers east of the Sandy River bridge. It is not enough to say only, "five kilometers from the center of town." You would not know whether this means 5 kilometers east or 5 kilometers west.

Similarly, if you wish to describe the position of a point on a line that you have drawn, you must specify some origin and state a distance and direction from that origin. But this time, the direction cannot be given as east or west, for the line may not run that way. You might try "right and left," but how would someone standing on the other side of the line interpret these directions? To get a description of direction along the line about which we can all agree, we shall call the line on one side of the origin positive, on the other side negative; we can then specify position on the line by a positive or negative number which gives both the distance (in some convenient units) and the direction of that point from the origin. We shall refer to such a number (with its sign and units) as the coordinate of the point. If we call the line the x coordinate line, we shall label these coordinates as x_1, x_2, x_3, etc. (Fig. 11 – 1).

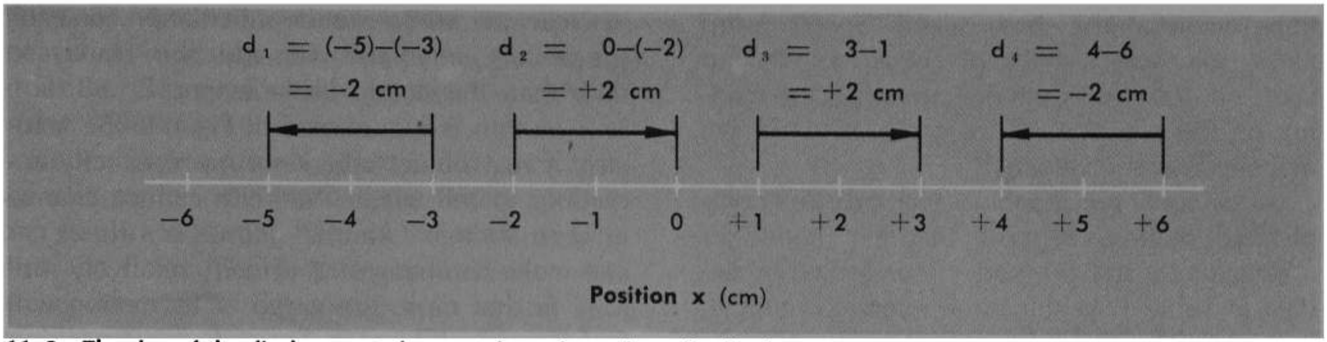


11–2 A position-time graph for motion along the x coordinate line.

To describe the motion of an object along the coordinate line, it is often convenient to make a graph of position against time. In such graphs, we usually plot the time along the horizontal axis and the position along the vertical axis. Figure 11 – 2 is an example of such a graph. There are many qualitative features about the motion which you can learn immediately from the graph.

The object was at position $x = 1.0$ cm at the time chosen as zero time. We are not given anything about its previous history. Between $t = 0$ and $t = 1.5$ sec, we see that the object is moving toward the point $x = 2.0$ cm and arrives there at $t = 1.5$ sec. Then it reverses its direction of motion and heads toward smaller numbers on the line. At $t = 2.5$ sec the object reaches the point $x = 0.5$ cm and there it remains at rest until $t = 3.5$ sec. Finally it moves on, passes the origin at $t = 4.2$ sec, and goes to $x = -1.0$ cm. There the graph stops and our information ends.

We shall often want to refer to the change of position in our study of motion, and we shall give it a special name, the *displacement*. If an object moves from position x_1 to position x_2, the displacement d is given by $x_2 - x_1$. Displacement can be either positive or negative (positive when x_2 is greater than x_1, negative when x_2 is less than x_1). As Fig. 11 – 3 shows, whether the displacement d is positive or negative depends only on the direction of the change of position; it does not depend on which side of the origin of the x coordinate line the displacement takes place. The arrows in the figure show the direction of the displacements. The displacements d_2 and d_3 are in the positive direc-



11–3 The sign of the displacement does not depend on where the displacement takes place, but only on its direction.

tion; they are positive in value and are exactly equivalent. The displacements in the negative direction, d_1 and d_4, are negative in value and exactly equivalent.

11–2 Velocity

Suppose you move steadily along a straight path. In order to predict where you will be at some future time, you need to know your present position, how fast you are going, and your direction of motion. The quantity which tells how fast and in which direction something moves we call the *velocity*.

We shall define the velocity of an object moving steadily as the displacement per unit time, or, in algebraic language

$$v = \frac{x_2 - x_1}{t_2 - t_1} = \frac{d}{t}$$

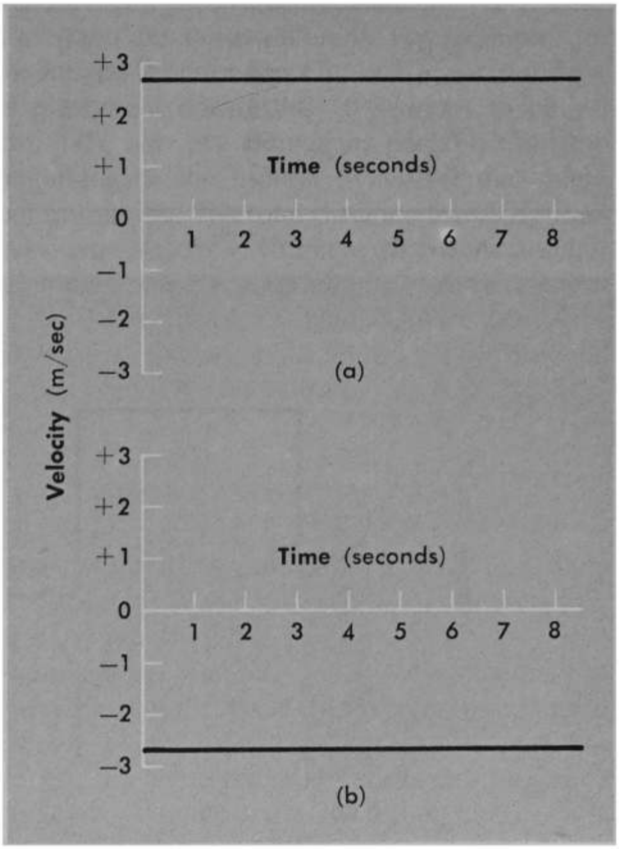
where $t = t_2 - t_1$ is the time interval during which the displacement occurred. As you can see, since time is always running forward, t is positive, so the sign of the velocity is determined only by the sign of the displacement; when the displacement is positive (toward the positive direction) the motion is in the positive direction and the velocity is positive, regardless of where it takes place on the coordinate line. Conversely, when the displacement is negative, the velocity is negative and the motion is in the negative direction.

Consider a simple example. Suppose that each of the displacements in Fig. 11 – 3 takes place in 0.4 sec. Then, representing the velocities as v_1, v_2, v_3, and v_4 for each of the displacements d_1, d_2, d_3, and d_4, we have from $v = \frac{d}{t}$

$$v_1 = \frac{-2\text{ cm}}{0.4\text{ sec}} = -5\text{ cm/sec}$$

$$v_2 = \frac{+2\text{ cm}}{0.4\text{ sec}} = +5\text{ cm/sec, etc.}$$

It is often convenient to represent the velocity of a moving object as a function of time on a graph. Figure 11 – 4 (a) shows such a graph for an object moving at constant velocity. Figure 11 – 4 (b) is for a similar motion but in the opposite direction. We say that both of these mo-



11–4 The velocity of an object moving steadily is graphed as a horizontal line on a velocity-time graph: (a) motion in the direction of increasing positive coordinate values; (b) motion at the same speed but in the opposite direction.

tions occur at the same speed. *Speed* designates only how fast the object is moving and takes no account of the direction. It is the reading on the speedometer of a car. It is just the magnitude of the velocity.

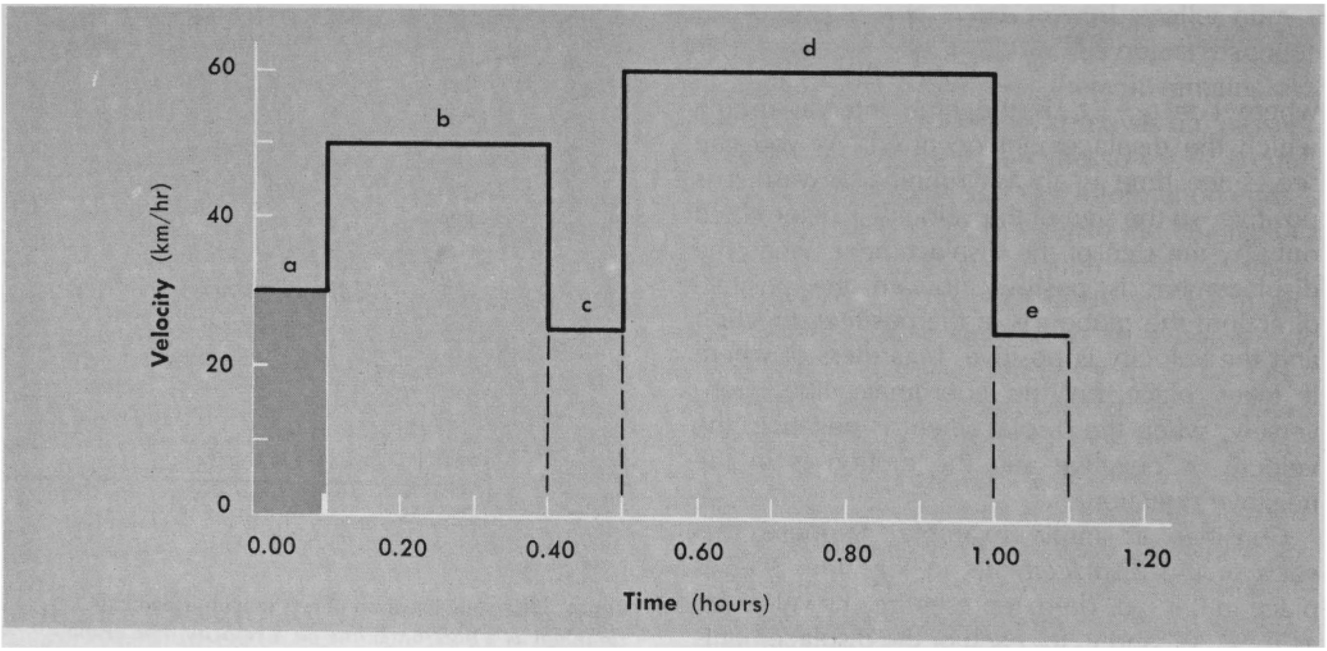
Let us apply our graphical methods to a complicated motion. Suppose that the motion of a car along a straight road is represented by the data in Table 1. Within each interval of time *a, b, c, d,* and *e,* the car travels at constant velocity; however, these velocities are different for different intervals. To compute the displacement during the first time interval (0.10 hr long), we use the equation $d = vt$. The result is 3.0 km.

TABLE 1 *Motion of a Car with a Variable Velocity*

TIME INTERVAL NUMBER	DURATION OF INTERVAL (hr)	VELOCITY DURING INTERVAL (km/hr)	DISPLACEMENT (km)
a	0.10	30	3.0
b	0.30	50	15
c	0.10	25	2.5
d	0.50	60	30
e	0.10	25	2.5

We can perform a similar calculation for each succeeding interval, and add the results to find that the total displacement is 53 km. This motion is represented in Fig. 11 – 5. Actually, a real object could not move exactly according to this graph. Velocity cannot change in such sudden "jumps." However, a real car can make its changes of velocity relatively rapidly. In that case, the graph of its motion will look very much like Fig. 11 – 5. In this discussion we shall ignore the impossibility of sudden jumps, so that we can keep our graph simple.

One great convenience of graphical presentation is that it enables us to see quickly when the car is going fast and when it is going slowly, in addition to indicating the direction of the motion. Thus, the higher positive velocities occur "high" on the graph of velocity versus time. Can the graph also tell us the displacement in each interval? The answer is "yes." Let us see how. During any one of the five intervals the position changes by an amount given by the equation $d = vt$. In any interval, the height of the graph tells us the velocity during the interval, and the horizontal length gives us the time. Thus v times t is the height times the base, or the "area" of the rectangle. This area is shaded for the first interval in Fig. 11 – 5. The units of these "areas" are different from the more common



11–5 The motion of a car moving at different velocities during different time intervals. The displacement in any interval is measured by the area enclosed between the graph and the horizontal axis.

cm^2 or in.^{2} because one side of the rectangle is measured in hours, and the other is measured in km/hr. The product in this case has units of hours × km/hr = kilometers.

Since heights and horizontal distances that we plot on the velocity and the time scales of our graph are proportional to the actual velocities and times involved, any two areas on the graph are exactly proportional to actual displacements of the car. This fact often allows us to decide at a glance in which time interval the greatest displacement occurs. For example, we can see that the area of the rectangle marked *d* in Fig. 11 – 5 is greater than that of any of the other rectangles. Therefore, we know without calculations that the position changes more during the interval *d* than in any of the other intervals.

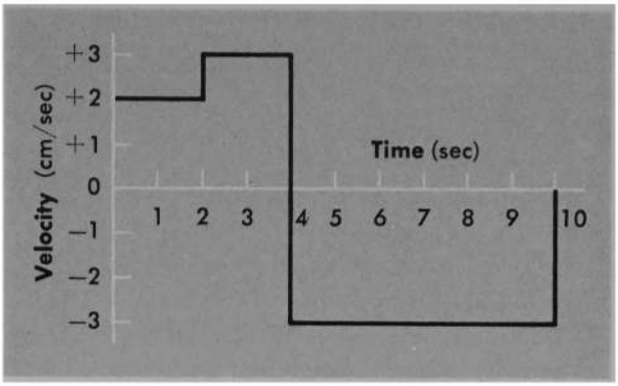
The total displacement of the car in 1.10 hr is obtained by adding up the "areas" of all the intervals in Fig. 11 – 5.

What if we have a velocity-time graph like that in Fig. 11 – 6, in which both positive and negative velocities occur? How do we find the displacement from the areas of the rectangles? During the first four seconds, the displacement is the sum of the areas of the first two rectangles: it is (+2 cm/sec)(2 sec) + (+3 cm/sec)(2 sec) = +10 cm. During the next 6 seconds the object had a negative velocity, back toward its starting point, and its displacement was negative. When we calculate the area of the third rectangle, it comes out negative: (−3 cm/sec) (6 sec) = −18 cm, representing a negative displacement. The total displacement between 0 sec and 10 sec that we find from the graph is, therefore,

$$d = (+10 \text{ cm}) + (-18 \text{ cm}) = -8 \text{ cm}$$

We can summarize the motion as follows: the object moved 10 cm in the positive direction during the first 4 seconds and then reversed its direction of motion, traveling 18 cm in the negative direction during the next 6 seconds. It ended up displaced −8 cm from its starting point.

We have seen that a velocity-time graph gives us information from which we can find the displacement during a time interval, but it does not tell us anything about the position of the object that is moving. To predict where an object will be at the end of the time interval, we also need to know its position at the beginning of the interval. The final position then is just the sum of the initial position and the change of position, that is, the sum of the initial position and the displacement.

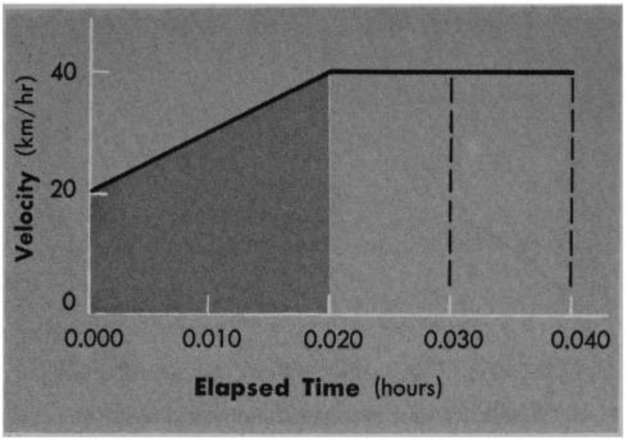


11–6 A velocity-time graph with both positive and negative velocities.

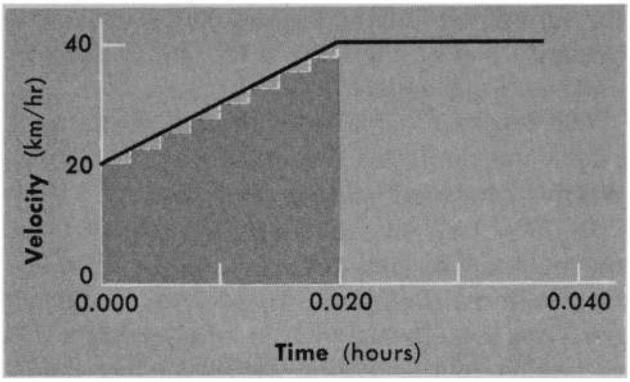
We must be careful not to confuse the displacement during a time interval with the total distance traveled during that interval. They can often be quite different. For example, for the motion described by Fig. 11 – 6, we found the displacement during the 10 seconds to be −8 cm. This was the change in position from the beginning of the motion (wherever that might be) to the end. The total distance traveled, however, was 10 cm + 18 cm = 28 cm. As another example: suppose you drive from San Francisco to New York and back again. The total distance traveled would be about 9500 km, but the total displacement would be zero!

11–3 VARYING VELOCITY

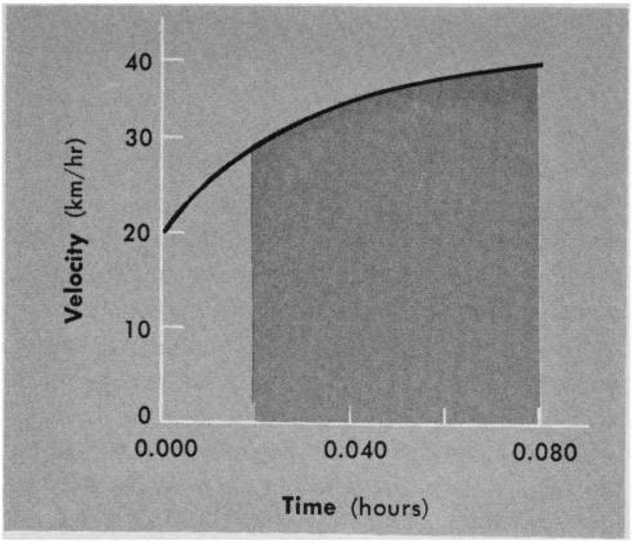
Now we shall use our graphical ideas to help analyze a more difficult problem. Figure 11 – 7 gives a graph of the velocity of a car versus elapsed time. Can we find the displacement of the car in the first 0.020 hr? We can try to multiply the velocity by the time, but we get into trouble, for we must now choose from a whole range of velocities. On the other hand, using the area under the graph, which worked as an alternate method for motion at constant velocity, might also serve here, allowing us to solve graphically a problem that presents difficulties when tackled algebraically.



11–7 The velocity-time graph for a car which is changing velocity during part of its trip. Does the shaded area give the change of position during the time interval from 0.000 to 0.020 hour?



11–8 In this figure, an imaginary car changes its velocity abruptly in steps, never moving faster than the real car of Fig. 11–7. If the steps are frequent enough, the shaded area is almost equal to the shaded area of Fig. 11–7.



11–9 In general, the displacement is given by the area under the velocity-time graph, no matter how the velocity changes.

Using the area to find the displacement looks reasonable, because we can approximate the sloping graph of Fig. 11 – 7 closely by the "stepped graph" of Fig. 11 – 8. This graph represents the motion of an imaginary car which starts at the same position as the real car, but which has a velocity that changes by a series of steps. At the beginning of each step, its velocity is equal to the velocity of the real car. While the imaginary car's velocity remains constant, the velocity of the real car increases, so the real car gets farther and farther ahead of the imaginary car. At the end of the step the imaginary car's velocity suddenly increases to the value of the real car's velocity at that instant. The imaginary car can never catch up with the real car, however, because the real car is going faster at almost every instant. But if we make the steps much smaller and more frequent, the two cars will never differ much in velocity, and hence the real car will never get far ahead. Thus, the shaded area, which, as we know from Section 11 – 2, gives the displacement of the imaginary car during the time interval, will also very nearly give the displacement of the real car during that interval. And this shaded area, when we use many steps, is practically the shaded area under the graph of Fig. 11 – 7 for the real car. Thus, we see that the displacement of the real car is given by the area under its velocity-time graph.

Since the area under the slant line of Fig. 11 – 7 is that of a trapezoid with a base of 0.020 hr and the two heights +20 km/hr on the left and +40 km/hr on the right, we can now find the displacement of the real car. The area of the trapezoid gives the displacement

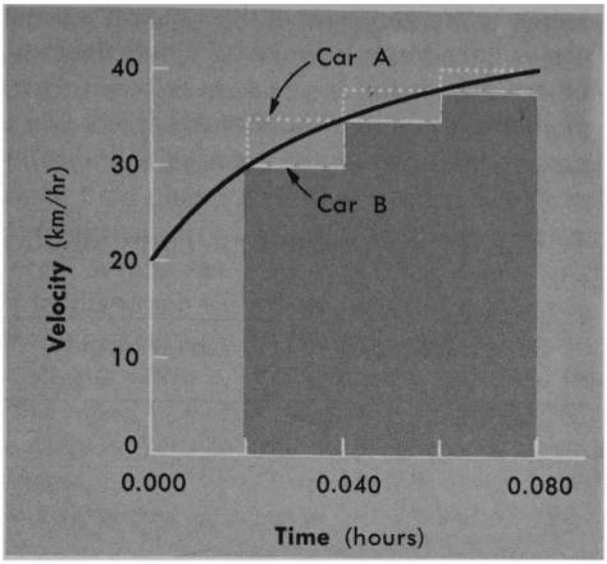
$$d = \frac{(20 \text{ km/hr} + 40 \text{ km/hr})}{2} (0.020 \text{ hr})$$

$$= 0.60 \text{ km}$$

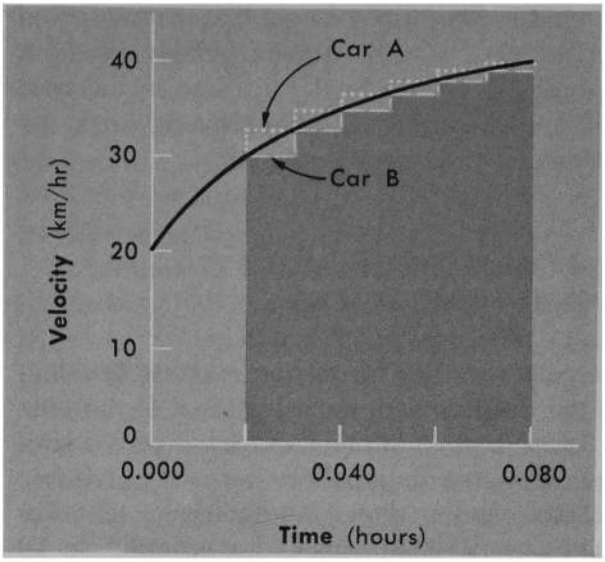
(You get the same answer by breaking the trapezoid up into a rectangle and a triangle and adding their areas.)

In general, even for more complicated velocity-time graphs, such as that of Fig. 11 – 9, the displacement is still given by the area under the graph. For example, for the time interval 0.020 hr to 0.080 hr, the displacement is given by the shaded area of Fig. 11 – 9. A proof of this is given in the box.

Even when we are not able to compute the



11–10 We can "bracket" the displacement of the car in Fig. 11–9 by imagining two other cars, A and B, that travel with different velocities as shown in this figure.



11–11 If the two cars of Fig. 11–10 change their velocities more frequently, they approximate the motion of the real car more closely.

Displacement as the Area under the Velocity-Time Graph

Here is an argument to show that the displacement during a time interval is always the area under the velocity-time graph, no matter what its shape. We can bracket the displacement of a real car by having two imaginary cars that change their velocity in large steps, as in Fig. 11 – 10. Car *B* at the initial time 0.020 hr has the velocity of the real car. For the first 0.020-hr period after this, it travels at 30 km/hr, while the real car speeds up from 30 km/hr to 35 km/hr; for the next 0.020-hr period, *B* travels at 35 km/hr; for the last 0.020-hr period, *B* travels at 38 km/hr. Meanwhile, *A* runs at 35 km/hr for the first 0.020-hr period after the initial time; during the second period it runs at 38 km/hr; and during the last period its speed is 40 km/hr. In a total time of 0.060 hr, *B* travels 0.60 km + 0.70 km + 0.76 km or 2.06 km altogether. Meanwhile, *A* travels 0.70 km + 0.76 km + 0.80 km = 2.26 km altogether. The displacement of the real car must lie somewhere between these two values, since it never moves faster than *A* or slower than *B*. Thus, we have bracketed the real car's displacement between the limits 2.06 km and 2.26 km.

Now make the steps smaller and more frequent, as in Fig. 11 – 11. If you calculate the displacement of *B* and *A* in the same 0.060-hr interval by finding the area under the graphs, you will find they are 2.12 km and 2.22 km respectively. This gives a smaller difference between the upper and lower limits than we obtained before. We can continue to decrease the difference between the upper and lower limits for the displacement by making the cars change velocity in shorter and shorter time intervals. The area representing the upper limit of the displacement and the area representing the lower limit become more and more nearly the same, and both of these areas get nearer and nearer to the area under the sloping curve. Mentally continuing this process to a very great number of very small steps, we prove that the "area" under the velocity-time graph does give the displacement of the original car.

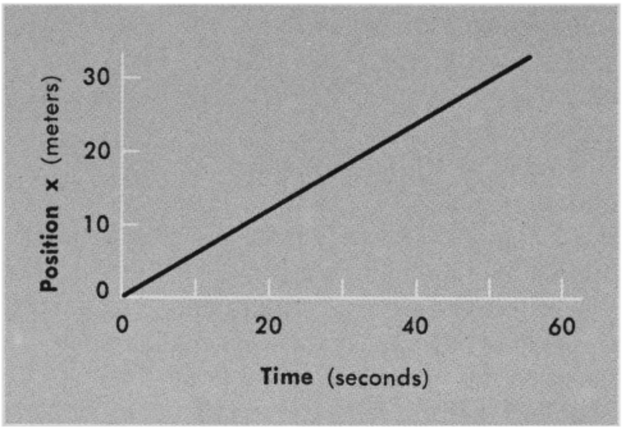
area from a formula, we can arrive at an approximate answer for the displacement in other ways. For instance, we can often approximate the motion fairly well by a collection of narrow rectangles, similar to the approximation we used in Fig. 11 – 8, and then calculate the area of the collection of rectangles. Another way is to divide the area into small squares, count the squares under the graph (making approximations when you come to only parts of squares), and then multiply the number of squares counted by the displacement that is represented by the area of one square. Either of these methods, or the method suggested in the box, will give you a very good approximation to the displacement during the interval.

11–4 ANALYSIS OF POSITION-TIME GRAPHS

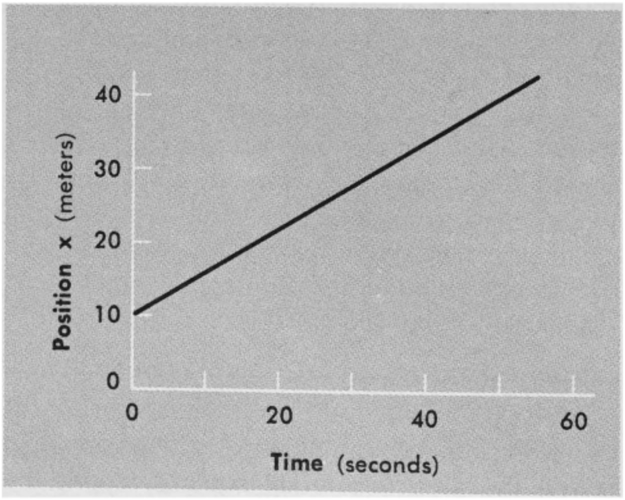
We have seen that for motion at constant velocity the displacement is proportional to the time of travel: $d = vt$. In other words, since v is constant, the area under the v versus t curve is a rectangle and is directly proportional to t. For a velocity of +0.60 m/sec, for example, in 10 sec the displacement is +6 m; in 20 sec, +12 m; and so on. This information is presented in Table 2. We can present this data on a position versus time graph if we know where the moving object was at the time taken as $t = 0$. For the moment, let us suppose that the object was just passing the origin of the coordinate line at $t = 0$.

TABLE 2 *Displacement-Time Relations for a Steady Velocity*

ELAPSED TIME (sec)	DISPLACEMENT (m)
10	6
20	12
30	18
40	24
50	30



11–12 The position-time graph for a steady velocity is a straight line.



11–13 The velocity is the same as in Fig. 11–12, but the initial position is different.

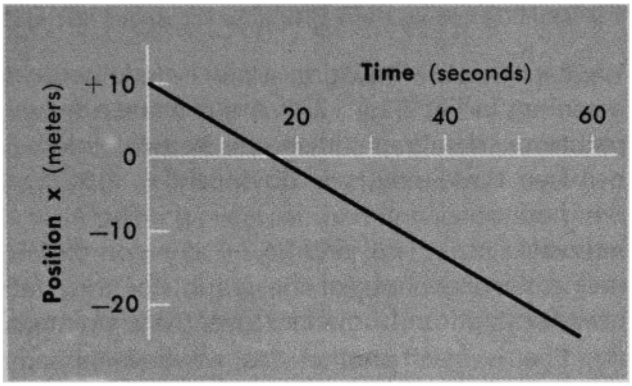
Then the graph of position versus time is shown in Fig. 11 – 12. Like all direct proportions, it is a straight line. What makes this line correspond to +0.60 m/sec is its steepness. How steeply the line rises depends on the velocity but not on the initial position. For example, if the velocity were greater, say +0.80 m/sec, the straight line would be steeper; it would rise the same amount in a shorter time interval. In Fig. 11 – 13, on the other hand, is shown the position-time graph of an object which also has a velocity of +0.60 m/sec, but which has an initial position of +10 meters at the time $t = 0$. This second object moves along the x coordinate line at the same velocity as the object of Fig. 11 – 12. The two graphs have equal steepness. If they start their motions at the same time, the second object will always be just 10 m ahead of the first, never catching up and never falling behind.

In Fig. 11 – 14 we have a different situation. The graph starts at +10 meters at $t = 0$ sec and *falls* instead of rises at the rate of 6.0 meters every 10 seconds. Thus, at $t = 10$ sec, the position is +4 meters; at $t = 20$ sec, the position is −2 meters; and so on. This represents motion in the negative direction along the x coordinate line; the velocity is −0.60 m/sec.

Now let us study the situation of Fig. 11 – 15. How can we tell from its graph which object has the larger velocity? The answer is simple. One curve climbs more steeply than the other.

For a given time interval, the steeper curve gives a greater change of position. Since it is the object with the larger velocity that has a greater displacement in any given time interval, the larger velocity corresponds to the more steeply rising graph. Object *A*, therefore, has a larger velocity than object *B*. In fact, we can see from the graph that at $t = 40$ sec, object *A* overtakes object *B* and passes it.

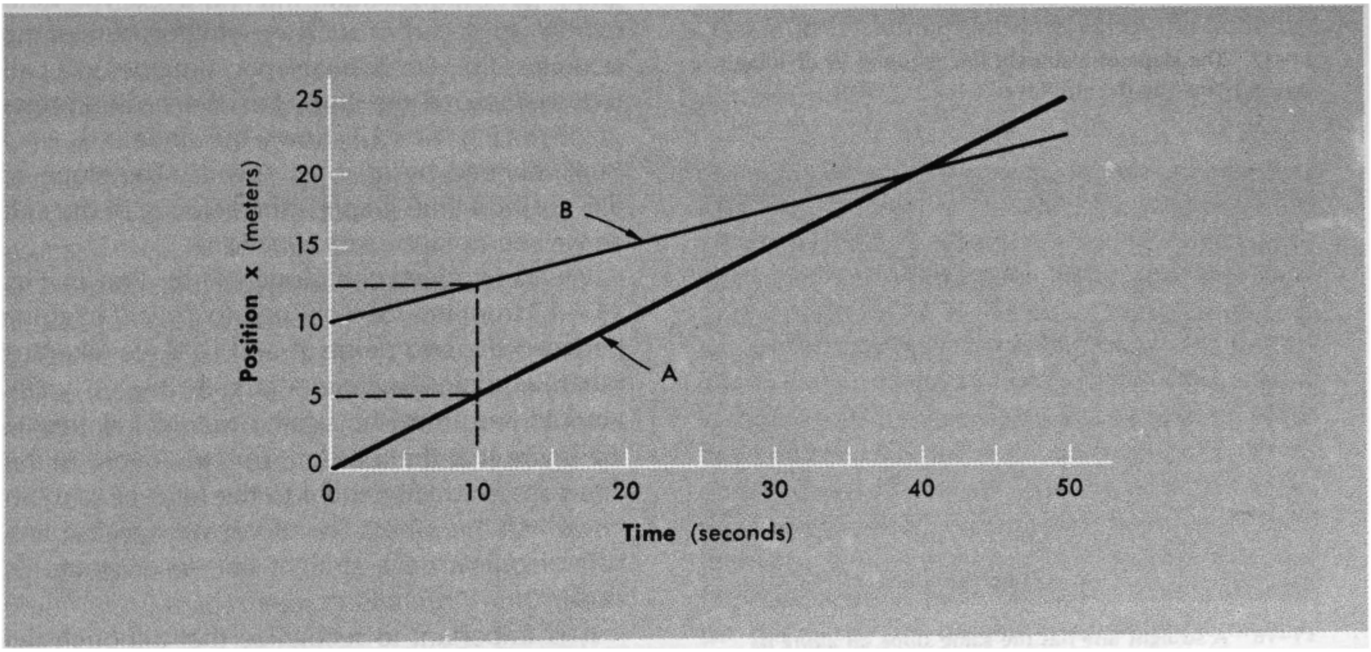
From the graph we can calculate the velocities of the two objects. In 10 sec, for example, *A* goes from the position $x_1 = 0.0$ m to $x_2 = 5.0$ m. Its displacement is +5.0 m and its velocity is therefore 0.50 m/sec. In the next 10 sec, it again changes position by 5.0 m, from $x_2 = 5.0$ m to $x_3 = 10.0$ m; its velocity is still 0.50 m/sec. Because the graph is a straight line, the displacement of *A* is the same for every 10 sec; therefore, the velocity of *A* is 0.50 m/sec all the time. Object *B* also has constant velocity. In each 10 sec its position changes by 2.5 m, from $x_1 = 10.0$ m to $x_2 = 12.5$ m in the first 10 sec; from 12.5 m to 15.0 m in the next 10 sec, and so on. Its velocity is therefore always 0.25 m/sec.



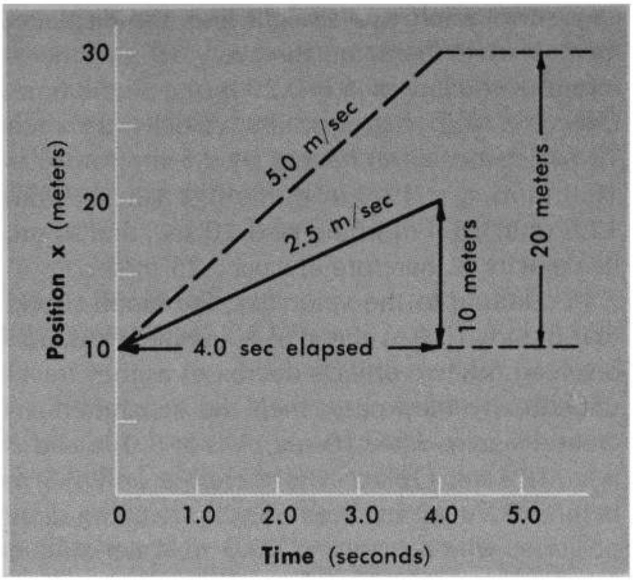
11–14 Motion in the negative direction along the coordinate line.

In addition to the velocities, the graph shows that *B* starts 10.0 m ahead of *A*, but the separation between the two objects decreases as they travel until finally they pass; then the separation increases again. After 10 sec, *A* is at 5.0 m and *B* is at 12.5 m. Object *A* is therefore only 7.5 m behind *B*. At 40 sec both objects are at the same position, which is at $x = 20.0$ m. Later still, at 50 sec, object *A* is now ahead of *B*; it is at 25.0 m, while *B* is only at 22.5 m.

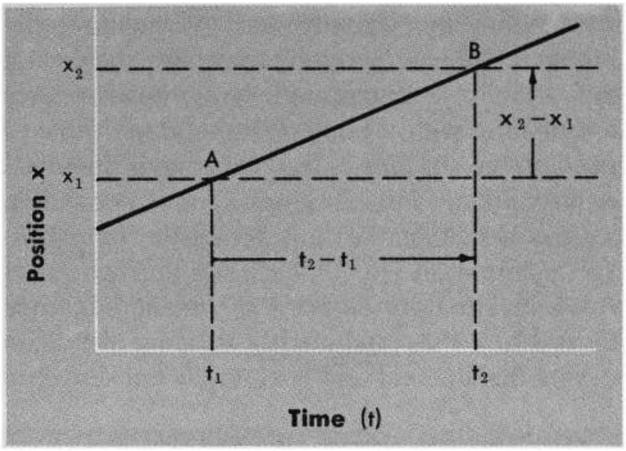
In Section 11 – 2 we saw that on a graph of velocity versus time we could tell at a glance at what times the velocity was the greatest; the higher the line occurred on the graph, the greater the velocity it represented. Now, however, we are dealing with a quite different graph—that of *position* versus time. The velocity is involved only indirectly in such graphs and is *not* shown by the height of the line above the time axis. For example, in Fig. 11 – 15, the line for object *B* is above that for object *A* in the entire interval from 0 to 40 sec, although *B* is being overtaken during this interval and is certainly the slower of



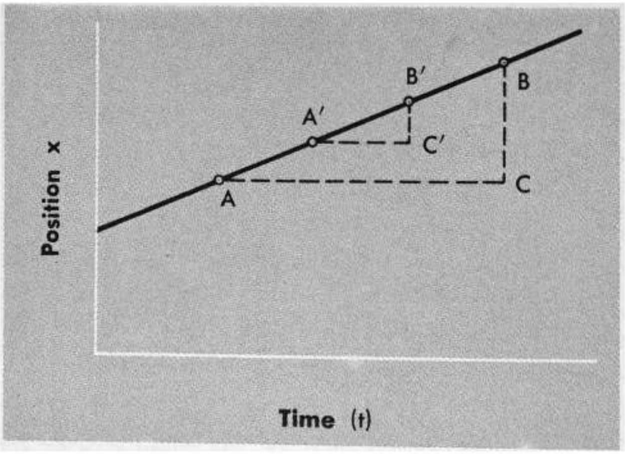
11–15 At what time will object A overtake object B?



11–16 Larger velocities give steeper graphs of position versus time.



11–17 The slope of a straight line is found by dividing (x_2-x_1) "up" by (t_2-t_1) "over."



11–18 A straight line has the same slope all along its length.

the two. On a position versus time graph, it is the steepness alone that tells the velocity.

Figure 11 – 16 emphasizes the relationship between steepness and velocity. The solid line is drawn for an object traveling at 2.5 m/sec. We see that it has traveled a distance of 10 m during the first 4.0 sec. An object going 5.0 m/sec travels 20 m during the 4.0-sec interval and is described by the steeper dashed line of the graph.

11–5 SLOPE

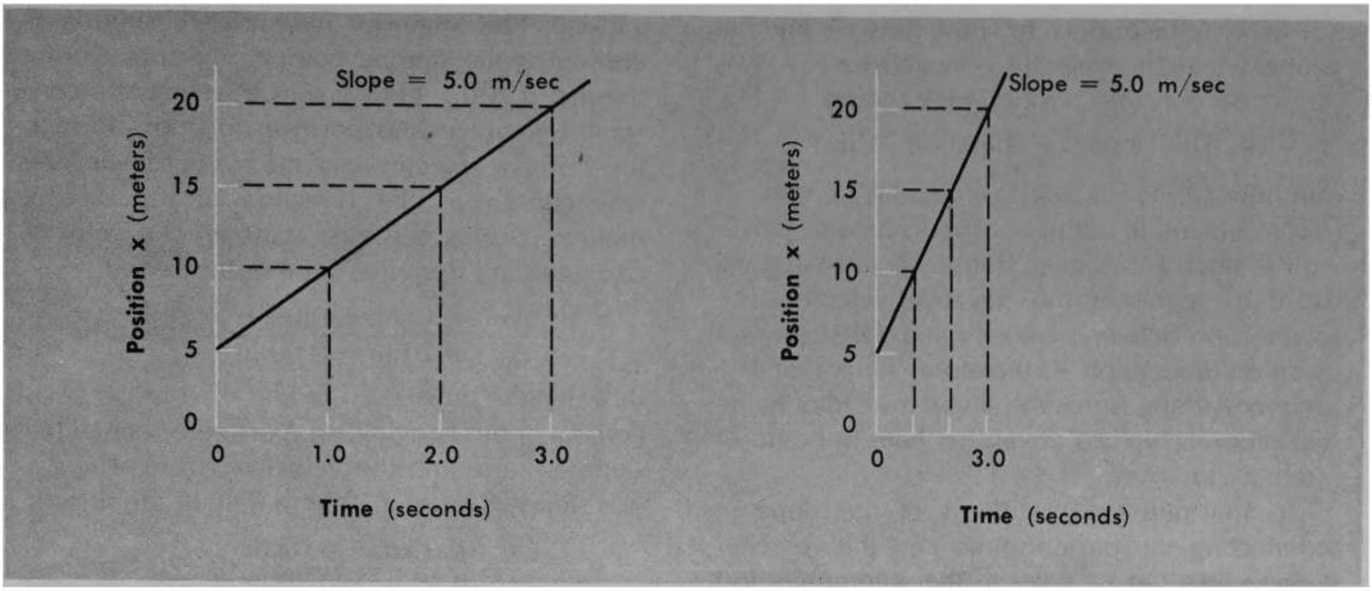
In general, if we pass at constant velocity from position x_1 at time t_1 to position x_2 at time t_2, the velocity v is, by definition

$$v = \frac{x_2 - x_1}{t_2 - t_1}$$

We can give this equation a precise geometrical meaning. In Fig. 11 – 17, if *A* and *B* are any two points on the graph, then the vertical interval between these points is obviously $x_2 - x_1$ and the horizontal interval is $t_2 - t_1$. These two intervals, $(x_2 - x_1)$ and $(t_2 - t_1)$, completely define the steepness of the graph, for they tell how far "up" and how far "over" one point on the line is from another. As we have already discovered, the velocity depends upon the steepness of the position-time graph; and the steepness of the graph depends upon how far up the graph goes for a certain interval over. The ratio of "up" to "over" is such a useful measure of the steepness that we give it a special name. We call it the *slope* of the line. For the position-time graph, as Fig. 11 – 17 shows, the slope is (x_2-x_1) "up" divided by $(t_2 - t_1)$ "over." The slope of the position-time graph is the velocity of the car, as we see from the last equation.

We determined the slope of the line in Fig. 11 – 17 from the ratio of "up" to "over" in going between the two points *A* and *B*. If we take any other two points such as *A′* and *B′* (Fig 11 – 18), we can see from the similar triangles dotted in the figure that the ratio of "up" to "over" in the one case is exactly equal to the ratio of "up" to "over" in the other. Therefore, we can use any pair of points on a straight line to compute its slope.

It is important to recognize that although the slope is related to the "steepness" of a graph,



11–19 The slopes of the two graphs are the same, although their appearances differ. They will appear the same only when plotted to the same scale or on the same graph.

the angle between the graphed line and the horizontal has no particular significance, since we could change the angle by replotting the data on a different scale, as shown in Fig. 11 – 19. It is only when we are comparing lines on the same graph, as in Fig. 11 – 15, or on graphs plotted to the same scale, that the angle between the line and the horizontal axis helps us compare slopes. In general, we must measure the vertical and horizontal intervals between two points on the graph and compute their ratio in the appropriate units — m/sec, for instance — as in Fig. 11 – 19.

Our discussion of slope applies to other graphs; nothing restricts us to the graph of distance versus time. We shall shortly see the significance of the slope of a graph of velocity versus time. In fact, we shall be dealing with slopes so often that it will be worthwhile to introduce a shorthand notation to indicate the process used in finding them. For instance, the slope of the position-time graph is always equal to the ratio of an interval of position $(x_2 - x_1)$ to an interval of time $(t_2 - t_1)$. The Greek letter delta, written as Δ, is often used as an abbreviation for the phrase "an interval of." The symbol Δ is a Greek capital D chosen to stand for "difference," or "change of," or "increase of," or "an interval of." Thus, Δx means "an interval of position" and Δt means "an interval of time." We read them "delta ex" and "delta tee."

We might compare the symbol Δ with some other algebraic symbol, for instance with "$\sqrt{}$," which means "take the square root of." In the expression $\sqrt{a}$, the symbol a stands for a number (or physical quantity) and $\sqrt{}$ tells us what to do with a. In a similar manner, in the expression Δt, the t stands for a physical quantity, and Δ tells us what to do with t. It says "take an interval" of t or "take the difference between two values" of t.

When the ratio of two intervals is involved, as it is in determining a velocity, it is customary to write the ratio as a fraction, with the understanding that the interval in the numerator "takes place during" the interval of the denominator.

Thus

$$v = \frac{\Delta x}{\Delta t}$$

which is read "vee equals delta ex over delta tee," means "to find the velocity, take the interval of position Δx in the time interval Δt during which the change of position took place and divide that position interval by that time interval."

In general, when we write $\Delta a/\Delta b$, we always mean that we shall use the interval of a that corresponds to a given interval of b. After all, we are interested only in the ratios of related intervals. Note that it makes no sense to separate the Δ from the a or b. The whole symbol Δa has a special meaning: an interval of a. It does not mean Δ multiplied by a.

You may be curious to know how we find the slope when the velocity is negative. This was the case for the position-time graph of Fig. 11 – 14. The slope, by definition, is just $\frac{x_2 - x_1}{t_2 - t_1}$ but now $x_2 - x_1$ is a negative quantity, a negative displacement. In our new symbols Δx is negative, so the slope is negative. But this is just what we want to represent the negative velocity. The connection between velocity and the slope of a position-time graph is, therefore, still valid. The only new thing is that the slope, instead of being the ratio of "up" to "over," is here the ratio of "down" to "over."

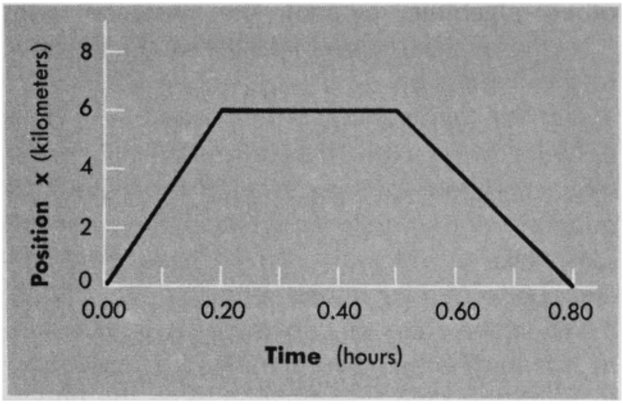
To illustrate the usefulness of the slope in connecting the position-time and the velocity-time graphs, let us analyze the automobile trip described by the position-time graph of Fig. 11 – 20. The origin for measuring position was chosen as the starting point of the trip. During the first 0.20 hr, the car was traveling at a constant velocity in the positive direction. In fact, the positive direction on the x coordinate line was chosen as the direction of the starting motion. During this first 0.20 hr, the velocity can be found from the slope of the curve:

$$v_1 = \frac{\Delta x}{\Delta t} = \frac{6.0\text{ km} - 0.0\text{ km}}{0.20\text{ hr} - 0.00\text{ hr}} = 30\text{ km/hr}$$

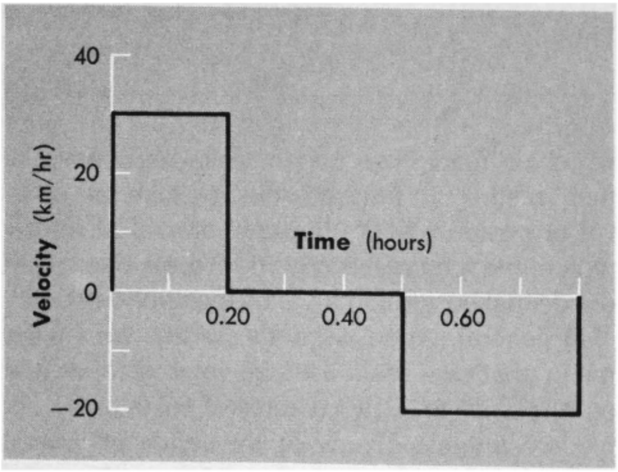
Between the times 0.20 hr and 0.50 hr the position was not changing, so Δx is zero; hence the slope is zero, and the velocity is zero. The car was stopped. From 0.50 hr to 0.80 hr the slope is

$$v_2 = \frac{\Delta x}{\Delta t} = \frac{0.0\text{ km} - 6.0\text{ km}}{0.80\text{ hr} - 0.50\text{ hr}} = -20\text{ km/hr}$$

The velocity of the car was negative, so it was headed back again toward the starting point. In fact, at $t = 0.80$ hr it reached the starting point again, so its displacement at that time was zero, even though it had traveled a total distance of 12.0 kilometers. We now have the data needed to plot a velocity-time graph for this motion. It is shown in Fig. 11 – 21. You will see that in this figure the positive area between 0.00 hr and 0.20 hr (which represents the positive displacement occurring during that time interval) is just equal to the negative area between 0.50 and 0.80 hr (which represents the negative displacement occurring during that time interval). Thus again we see that the total displacement of the trip was zero.



11–20 Position-time graph of a complete trip made by a car. Between which times was the car going the fastest?

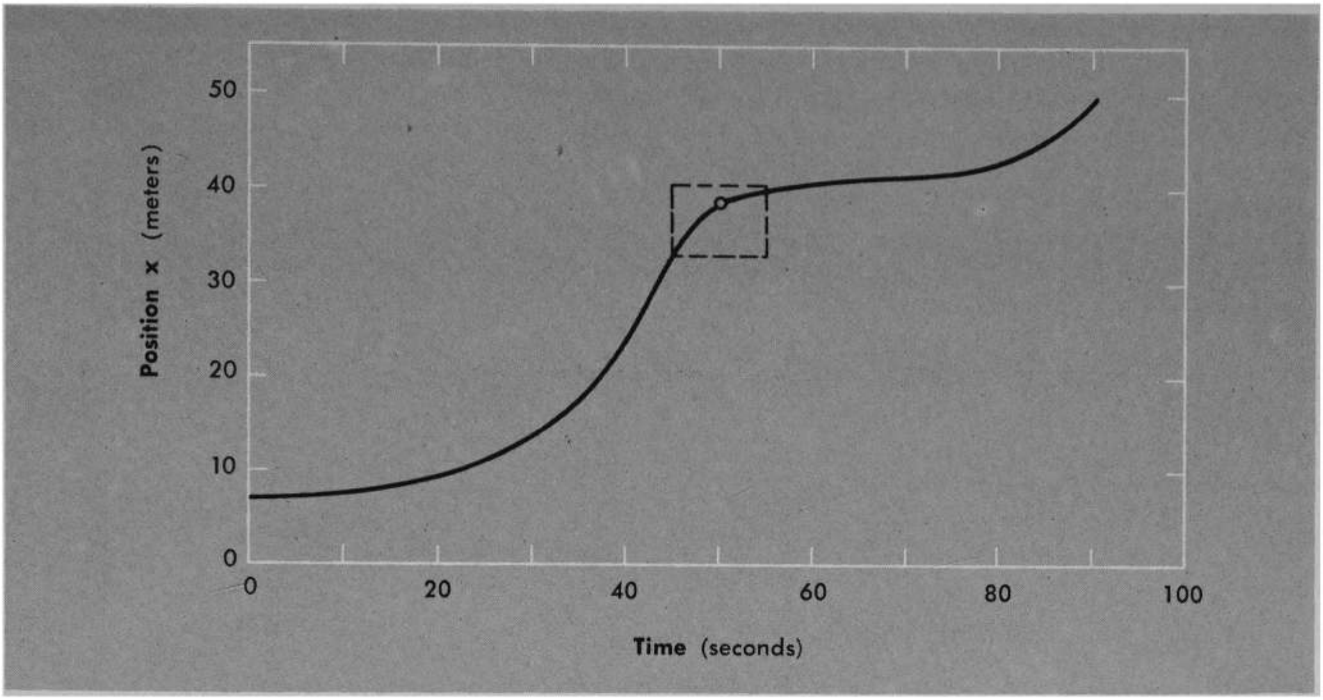


11–21 Velocity-time graph of the motion shown in Fig. 11–20. The portion of the graph below the horizontal axis indicates that the car is traveling in the direction opposite to its original direction.

11–6 INSTANTANEOUS VELOCITY —THE SLOPE OF THE TANGENT LINE

We have been considering motions in which the velocity was constant, either for the entire time or for different portions of the time. The position-time graphs we drew were therefore made of straight lines. Obviously, many motions do not occur at constant velocity. Figure 11 – 22 is a position-time graph of a more realistic situation in which the velocity is continually changing. How can we find the velocity of the object at any particular time?

Here there are no straight lines, and it is not obvious how to apply the methods we have used before. All that we can determine easily from



11–22 Position-time graph for an object with continually changing velocity.

the graph is the direction of the velocity. Since the position values are increasing with time, the object is traveling in the positive direction along the line; the slope is positive. (Of course, if the position values were decreasing as time increased, we would have a negative slope and the motion would be in the negative direction.) Our problem is to decide how to get the magnitude of the velocity from the graph. Let us consider the time 50 sec after the start of the motion.

Let us look with a magnifying glass at the part of the graph within the dashed rectangle. If we magnify it tenfold, we obtain the graph of Fig. 11 – 23. Note how much straighter the curve on this new figure looks, compared with that of Fig. 11 – 22. A still greater magnification of the dashed rectangle in Fig. 11 – 22 shows us the interval which covers only 0.5 sec before and after the 50-sec mark (Fig. 11 – 24). With this graph, we can tell the velocity by measuring the slope of the "straight" line. We choose two points near 50 sec, for example, 1 and 2 in Fig. 11 – 24; then reading from this graph, we find

$$t_1 = 49.86 \text{ sec} \qquad x_1 = 38.42 \text{ m}$$
$$t_2 = 50.16 \text{ sec} \qquad x_2 = 38.58 \text{ m}$$

Consequently, the slope is given by

$$\frac{\Delta x}{\Delta t} = \frac{x_2 - x_1}{t_2 - t_1} = \frac{+0.16 \text{ m}}{0.30 \text{ sec}} \approx +0.53 \text{ m/sec}$$

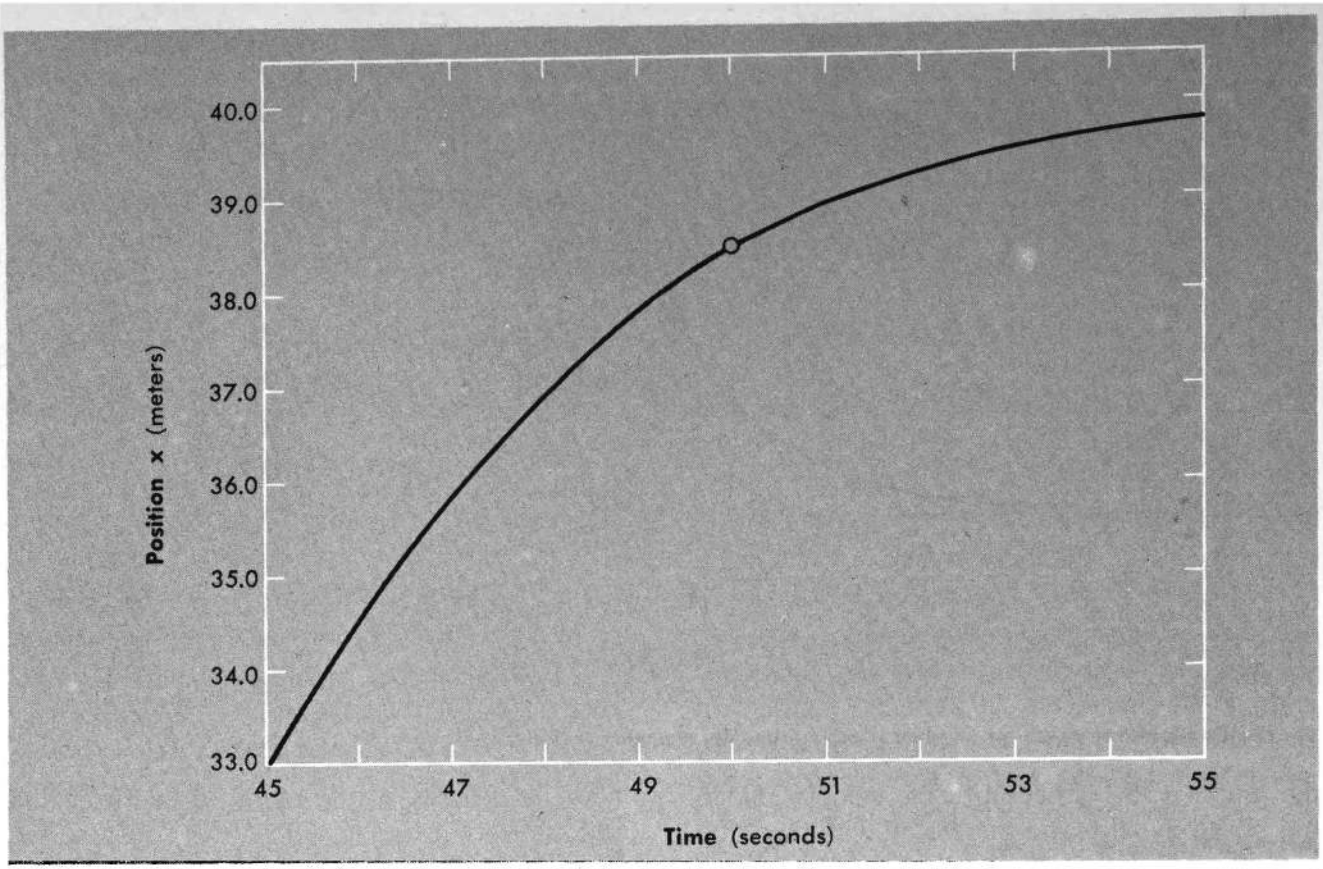
and the velocity at the point 50 sec from the start is very close to +0.53 m/sec. Thus, we can say that the velocity of the car at $t = 50$ sec is 0.53 m/sec in the positive direction.

The magnified part of a graph looks straighter than the whole graph because in the magnified picture we look at only a small portion of the unmagnified graph. When we magnify sufficiently, we look at only a small interval of x and t. In effect, therefore, we find the slope of a small portion of the curve by taking the ratio

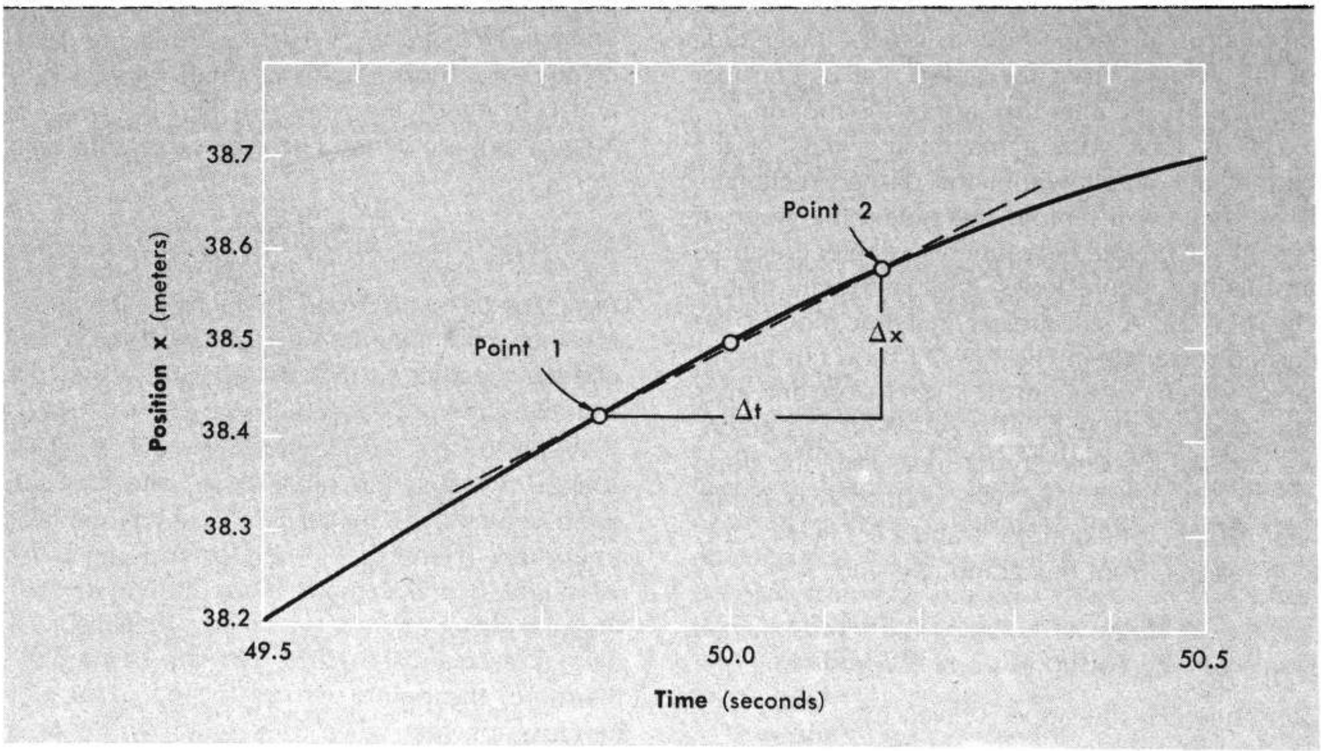
$$\frac{\Delta x}{\Delta t} = \frac{x_2 - x_1}{t_2 - t_1}$$

for a pair of points 1 and 2 which are very close together. The points we use must be close enough together so that the graph is essentially a straight line in between.

When the part of the graph we use is nearly straight, we find the same slope, and thus the same velocity, no matter which points we take to evaluate it. In Fig. 11 – 24, for example, if we let points 1 and 2 move along the curve until they are closer to the 50-sec mark, we get almost the same result as before. Try it yourself. Furthermore, the points we use to find $\Delta x/\Delta t$ will be close together when Δt is small; and if Δt is small enough, $\Delta x/\Delta t$ will have a definite value for a smooth position-time curve around a given



11–23 In this figure, the part of the graph enclosed in the dashed rectangle of Fig. 11–22 is enlarged ten times.



11–24 At a magnification of 100, a small portion of the graph within the dashed rectangle of Fig. 11–22 appears to be almost straight.

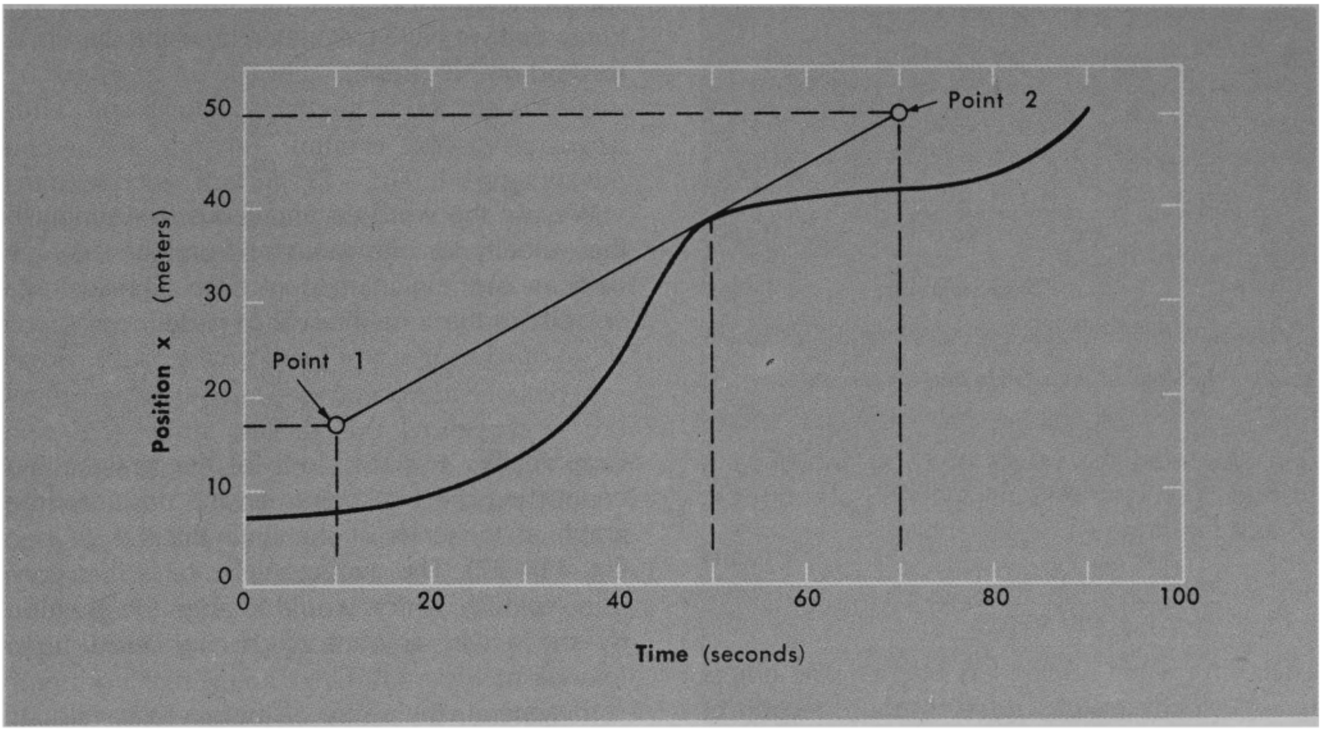
point. We therefore say that the velocity at a particular instant is given by the limit of $\Delta x/\Delta t$ as Δt "approaches zero," that is, as Δt gets smaller and smaller. Symbolically this statement is written

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}$$

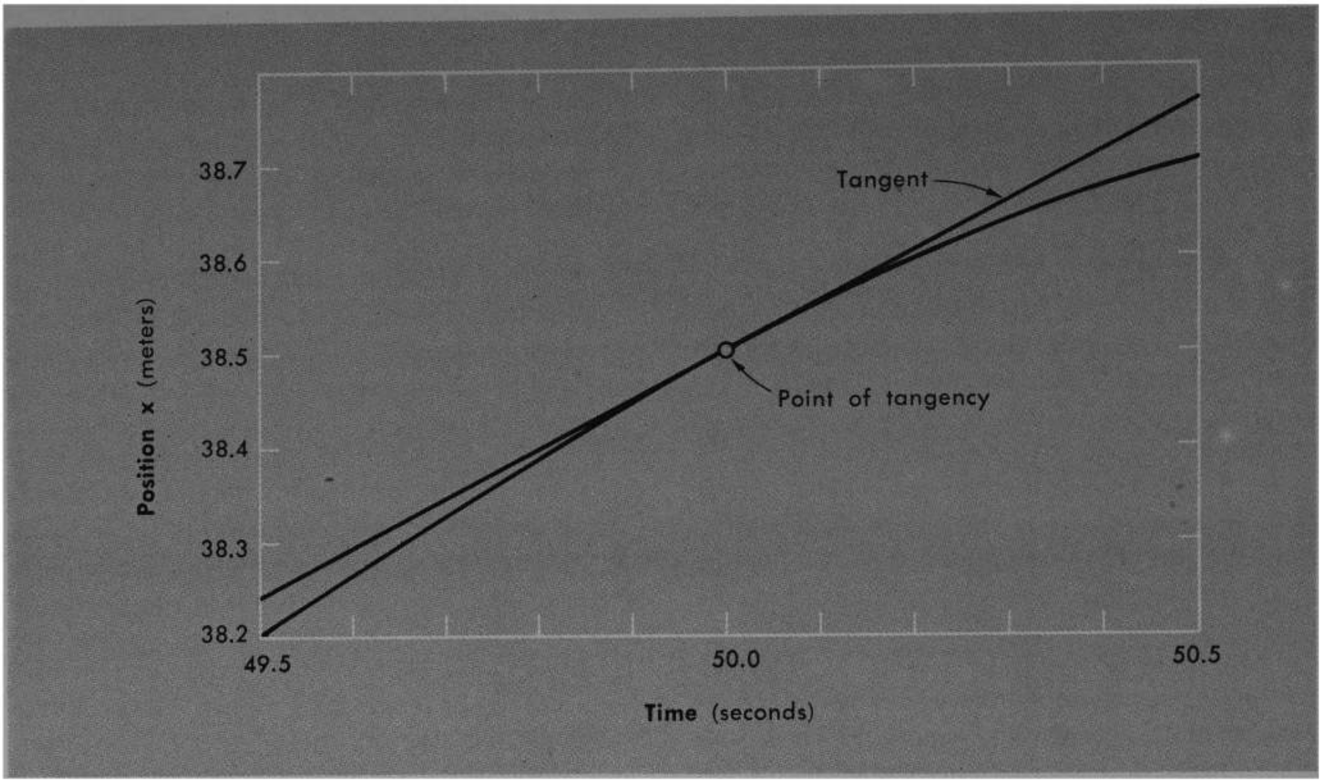
(We read "$\lim_{\Delta t \to 0}$" as "the limit as Δt approaches 0 of. . . .") The word "limit" here stands for the result obtained when we take t_1 and t_2 close enough together so that there is no noticeable change produced in the value of the ratio by using a still smaller interval between t_1 and t_2. We make the interval Δt shorter and shorter until "in the limit," as we say, we get a definite result. This procedure lies at the heart of the branch of mathematics called the calculus.

There may be special cases when this procedure is hard to apply or when it has no meaning. A plot of positions which changed suddenly from one straight-line slope to another would have no slope at the point of abrupt change; no velocity could be defined there. In fact, however, in all motions with which we have any direct experience, the velocity changes smoothly, even if sometimes rapidly. Therefore, the slope of the position plot does not change instantaneously; and we can take short enough time intervals to get the velocity at any given instant.

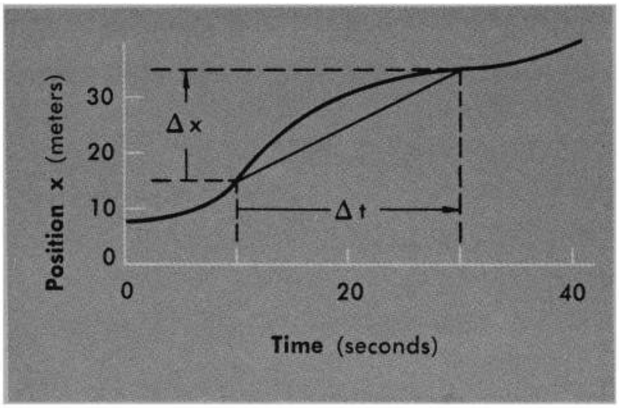
To use the procedure we have outlined, we need to have a magnified picture of the position-time graph—obtained, for example, by using a magnifying glass. Is there any simple way of getting the velocity without going through the steps of magnifying? The answer is "yes": draw a tangent and measure its slope. Suppose we drew a straight line tangent to the curve of Fig. 11 – 22 at the point for which $t = 50$ sec, as in Fig. 11 – 25. In this figure we can easily distinguish the tangent from the graph of x vs. t; but at hundredfold magnification as in Fig. 11 – 26, we see that the tangent line and the curve are hardly distinguishable over a small enough interval around the point of tangency. In this region they have the same slope. We can therefore use the slope of the tangent line to determine the slope of the graph at any particular point. Figure 11 – 25 shows us how we can use this method to get the speed of the car at $t = 50$ sec. First, we draw the tangent line to the curve at the point chosen. Then we choose two convenient points (labelled 1 and 2) on the tangent



11–25 To find the velocity at any instant, draw a line tangent to the curve at the point being considered. Then, by taking any two points on the tangent line, find the slope. The value of the slope is the velocity.



11–26 With 100-times magnification of Fig. 11–25, the curve and its tangent are indistinguishable near the point of tangency.



11–27 The slope of the straight line = $v_{av} = \Delta x/\Delta t$.

line. We read the values of x_1, x_2 and of t_1, t_2 to find $x_2 - x_1 = +32$ m and $t_2 - t_1 = 60$ sec, so that the slope is

$$\frac{x_2 - x_1}{t_2 - t_1} = \frac{+32 \text{ m}}{60 \text{ sec}} \approx +0.53 \text{ m/sec}$$

Therefore, when 50 sec has elapsed, the object has a velocity of 0.53 m/sec in the direction of increasing x.

The velocity at a given instant is called the *instantaneous velocity*. It is the magnitude of this instantaneous velocity that we read on a speedometer. Also it is the instantaneous velocity that we have calculated by using the slope method on our graphs:

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}$$

We use the word instantaneous to distinguish the velocity at one moment from the *average velocity* over an interval of time. The average velocity in the time interval Δt is defined as

$$v_{\text{ave}} = \frac{\Delta x}{\Delta t}$$

Graphically, it is the slope of the straight line connecting the two points on the position-time graph at the ends of the time interval Δt (see Fig. 11 – 27). The average velocity is that constant velocity which would change the position by the actual amount Δx in the actual time interval Δt.

In general, the value of the average velocity depends on which time interval Δt we are examining. In Fig. 11 – 25, the average velocity during the time interval 0 sec to 60 sec is not the same as the average velocity during the time interval 40 sec to 60 sec.

If, during Δt, the instantaneous velocity is changing noticeably, it may differ at almost every moment from the average velocity for that interval. At some moments it may be greater, at other moments smaller, than the average velocity. Only when the tangent to the position-time curve is parallel to the average-velocity straight line are they the same. For example, on the last graph they are the same only at $t = 18$ sec. The instantaneous velocity can be thought of as the average velocity over a time interval so short that the changes in velocity occurring during that interval may be neglected. We have already studied the case in which the instantaneous velocity equals the average velocity at every moment: motion at constant velocity. In this case, no matter whether we choose a long or a short time interval, the average velocity will be the same; hence, the instantaneous velocity cannot be different. You can see this graphically from Fig. 11 – 18 by imagining the points A' and B' very close together.

11–7 ACCELERATION

A driver might observe that his car can reach 50 km/hr in 10 seconds starting "from a standstill." He would be talking about the *acceleration* of his car, how fast its velocity changes. Let us suppose that Fig. 11 – 28 is the graph of the velocity versus time for the car. For simplicity we start with a straight-line graph; for a real car the graph might not be straight, but we shall consider more complicated graphs shortly. Figure 11 – 28 says that the car speeds up to 50 km/hr in 10 seconds, or 5.0 km/hr in each second. (We write this as 5.0 km/hr/sec or 5.0 km/hr-sec and read it "5.0 kilometers per hour per second.") The quantity 50 km/hr in 10 seconds or 5.0 kilometers per hour per second is a measure of the *acceleration*.

Figure 11 – 29 is the graph of the velocity versus time for another car. Notice that at "zero" time the car is going 10 km/hr and that 30 seconds later it is going 55 km/hr. The velocity has therefore increased 45 km/hr in 30 seconds. The acceleration is therefore given by $a = 45$ km/hr in 30 seconds or $a = 1.5$ km/hr-sec.

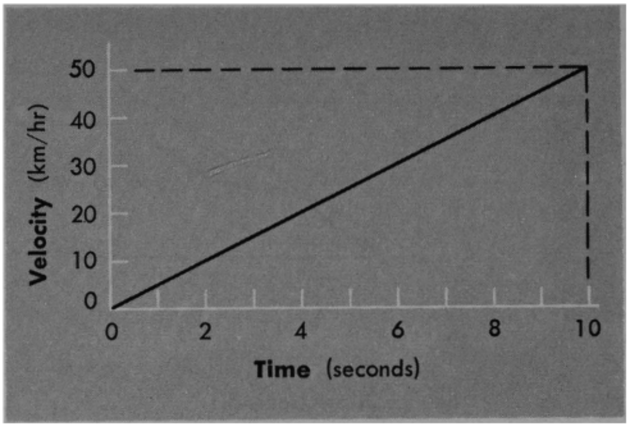
If we represent the quantities by symbols, we can use them in writing an algebraic expression for this calculation. Call the two times t_1 and t_2, and use v_1 and v_2 for the corresponding velocities at these two times. The acceleration is then given by the relation

$$a = \frac{v_2 - v_1}{t_2 - t_1} = \frac{\Delta v}{\Delta t}$$

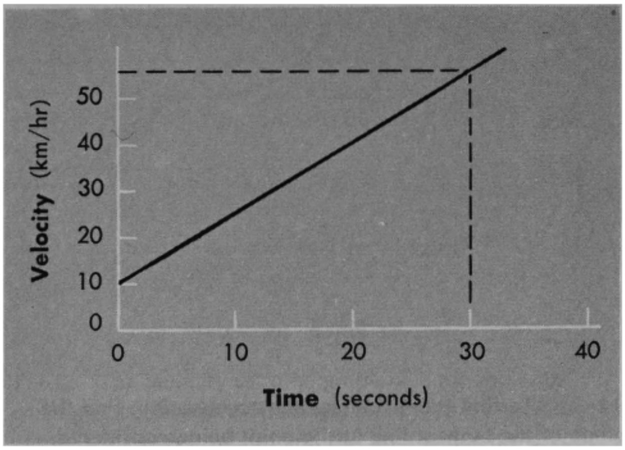
where we use the Δ notation that was introduced in Section 11 – 5.

There is a simple geometrical significance to this equation. Just as $\Delta x/\Delta t$ is the slope of the x versus t graph, so the right-hand side of this equation is the slope of the velocity versus time graph. So we can say that the acceleration is equal to the slope of the graph of v versus t.

If a graph of velocity vs. time is of the type shown in Fig. 11 – 30, we see that the car is slow-



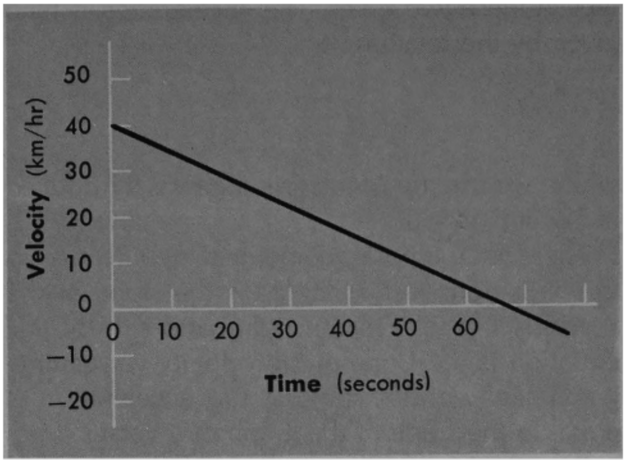
11–28 A velocity-time graph for an acceleration of 5.0 km/hr-sec.



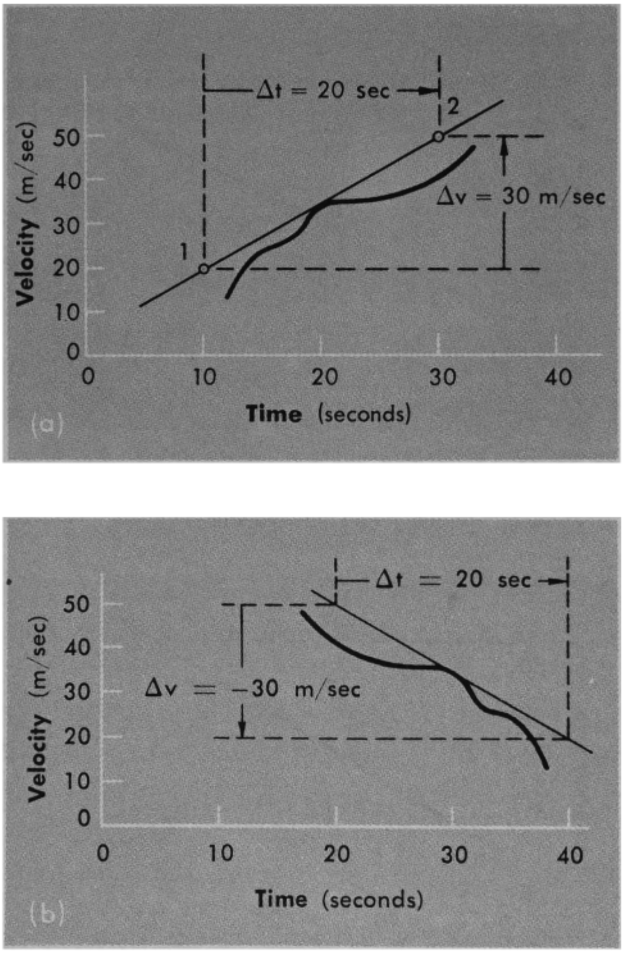
11–29 Graph of velocity vs. time for a car that is in motion at t = 0. What is the car's acceleration?

ing down. At the initial time $t = 0$, it is going 40 km/hr. Fifty seconds later it is going 10 km/hr. In this time interval, $\Delta t = 50$ sec, the change in velocity is $\Delta v = -30$ km/hr. Thus

$$a = \frac{-30 \text{ km/hr}}{50 \text{ sec}} = -0.60 \text{ km/hr-sec}$$



11–30 **Deceleration is shown by a negative slope on the velocity vs. time graph.**



11–31 **To find the instantaneous acceleration, find the slope of the tangent line at the point being considered. (a) Positive acceleration. (b) Negative acceleration.**

When $\Delta x/\Delta t$ has a minus sign, it means that the motion is backward, in the negative direction. What does a minus sign mean here for the acceleration? In this case the minus sign tells us that during the interval 0 to 67 sec, the car is moving in the positive direction but is slowing down; and from 67 sec to 80 sec it is moving in the negative direction and speeding up. In general, the minus sign means that the change in velocity Δv is in the negative direction. This happens either when an object which is moving in the positive direction along the x coordinate line slows down, or when an object which is moving in the negative direction speeds up. Indeed, if an object is moving in the positive direction but has a negative acceleration, it will slow down, stop momentarily, and then go backward, retracing its path at an increasing pace.

So far we have talked about constant acceleration—that is, about uniform changes of velocity. Let us now consider Fig. 11 – 31 (a), which shows another graph of velocity against time. What is the acceleration when 20 seconds have elapsed? In other words, what is the instantaneous acceleration at $t = 20$ seconds?

We know that if we magnified the portion near $t = 20$ seconds sufficiently, it would appear to be straight. The slope of this straight segment would then be the acceleration. Furthermore, as we have seen, instead of magnifying we can draw a straight line tangent to the graph at $t = 20$ seconds, and we can find the slope of the curve by measuring the slope of this tangent line. In view of the definition of the slope, the *instantaneous acceleration* at a time t equals the slope of the graph of velocity versus time. In Fig. 11 – 31 (a), by choosing two points (1 and 2) on the tangent line we find the instantaneous acceleration to be

$$a = \frac{v_2 - v_1}{t_2 - t_1} = \frac{50 \text{ m/sec} - 20 \text{ m/sec}}{30 \text{ sec} - 10 \text{ sec}}$$
$$= 1.5 \text{ m/sec}^2 \text{ at } t = 20 \text{ sec}$$

The notation m/sec^2 means "meters per second per second." In Fig. 11 – 31 (b), the instantaneous acceleration is

$$
\begin{aligned}
a &= \frac{20\ \text{m/sec} - 50\ \text{m/sec}}{40\ \text{sec} - 20\ \text{sec}} \\
&= -1.5\ \text{m/sec}^2 \text{ at } t = 30\ \text{sec}
\end{aligned}
$$

We can also express the instantaneous acceleration in the language of Section 11 – 6 as

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t}$$

In addition to the instantaneous acceleration at each moment, we define the *average acceleration* for any time interval Δt as

$$a_{\text{ave}} = \frac{\Delta v}{\Delta t}$$

This tells us the overall change of v in that particular time interval. The instantaneous acceleration at time t is the average acceleration for a very short time interval which includes t. If the acceleration does not change as time goes on, the average acceleration and the instantaneous acceleration are equal at any time.

11–8 USEFUL RESULTS FOR MOTION WITH CONSTANT ACCELERATION

When the acceleration is constant, the methods that we have been developing enable us to express several useful relationships in the form of equations. Consider an object whose velocity increases uniformly with time as shown by the graph in Fig. 11 – 32. Let us see what happens during a certain time interval. At the beginning of this time interval the velocity is called v_i (the subscript "i" means initial), and the corresponding time is called t_i. At the end of this time interval, the velocity is called v_f (the subscript "f" means final), and the corresponding time is called t_f. Because the acceleration is constant, we have

$$v_f - v_i = a(t_f - t_i) = at \qquad (1)$$

where t is the time interval between the initial time t_i and the final time t_f. To find the displacement during this interval we can compute the shaded area in Fig. 11 – 32. From the trapezoid formula (or by adding the "areas" of the rectangle and the triangle), we get

$$d = \tfrac{1}{2}(v_f + v_i)t \qquad (2)$$

a relationship between the displacement, the initial and final velocities, and the duration of the time interval.

We can also find a relation between the speeds and the displacement for constant acceleration. Solving Equation (1) for t, we get

$$t = \frac{v_f - v_i}{a}$$

Now, substituting this expression for t in Equation (2), we get

$$d = \tfrac{1}{2}(v_f + v_i)\frac{(v_f - v_i)}{a}$$

Finally, on carrying out the multiplication, we have

$$2ad = v_f^2 - v_i^2$$

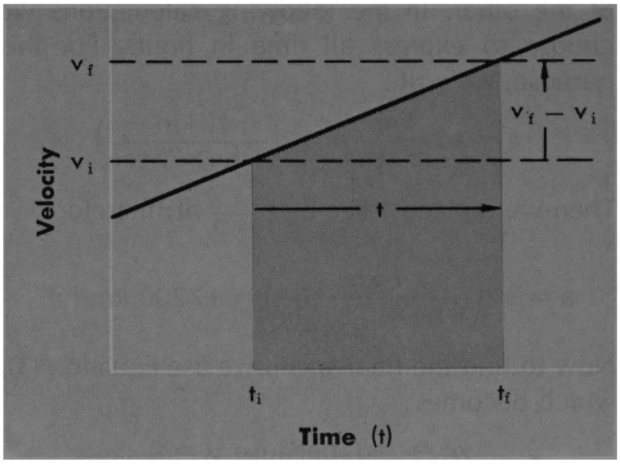
or

$$v_f^2 = v_i^2 + 2ad \qquad (3)$$

Because of the squares on the v_f and v_i, we lose our information about the directions (+ or −) of the velocities. Thus, we have only a relation between the squares of the speeds, the displacement, and the acceleration. However, it is often useful, and will even prove essential later on, in our discussion of kinetic energy.

Also, if we solve for the final velocity v_f in Equation (1) and substitute it into Equation (2), we get a relation connecting the displacement, the acceleration, the time interval, and the initial velocity:

$$d = v_i t + \tfrac{1}{2}at^2 \qquad (4)$$



11–32 The velocity vs. time graph of a motion with constant acceleration. The slope $(v_f - v_i)/(t_f - t_i)$ of the graph is the acceleration, and the area under the graph is the displacement.

There is a simple meaning to the two terms on the right-hand side of Equation (4). The term $v_i t$ gives the displacement of the car if its velocity had stayed constant at v_i during the time interval, while the second term gives the displacement the car would have if it had started from rest at the initial time. You can verify these statements by first setting $a = 0$, then v_i equal to zero, in Equation (4).

Some examples will show you the use of these equations. Suppose a sports car accelerates at a constant rate, changing its velocity in 6.0 seconds from +10 m/sec to +28 m/sec. What is the acceleration and how far does the car go in this time interval?

We find its acceleration from Equation (1):

$$a = \frac{v_f - v_i}{t_f - t_i} = \frac{28\text{ m/sec} - 10\text{ m/sec}}{6.0\text{ sec} - 0.0\text{ sec}}$$
$$= +3.0\text{ m/sec}^2$$

We find the displacement from Equation (4):

$$d = (10\text{ m/sec})(6.0\text{ sec}) + \tfrac{1}{2}(3.0\text{ m/sec}^2)(36\text{ sec}^2) = +114\text{m}$$

You can check this result for d by using Equation (3).

As another example, suppose a car starts from rest and accelerates at +2 km/hr-sec for a distance of $\frac{1}{9}$ km. How fast will it then be traveling and how long does it take to cover this distance?

We have two units of time, hours and seconds. It is convenient to eliminate one in favor of the other. In the following calculations we choose to express all time in hours. For this purpose, we write

$$a = +2\frac{\text{km}}{\text{hr-sec}} = +2\left(\frac{1\text{ km}}{1\text{ hr} \times 1\text{ sec}}\right)$$

Then we replace 1 sec by $\frac{1}{3600}$ hr; therefore,

$$a = +2\left(\frac{1\text{ km}}{1\text{ hr} \times \frac{1}{3600}\text{ hr}}\right) = +7200\text{ km/hr}^2$$

Now to find the final speed we use Equation (3), which becomes

$$v_f^2 = 2ad, \text{ because } v_i = 0$$

The result is

$$v_f^2 = 2\left(7200\frac{\text{km}}{\text{hr}^2}\right)\left(\tfrac{1}{9}\text{ km}\right) = 1600\text{ km}^2/\text{hr}^2$$

$$v_f = 40\text{ km/hr}$$

The direction of v_f will be the same as the direction of a, since the object started from rest. Thus, the final velocity is +40 km/hr. We can find the time from Equation (1), which becomes $v_f = at_f$, since v_i and $t_i = 0$. Therefore

$$t_f = \frac{+40\text{ km/hr}}{+7200\text{ km/hr}^2} = \tfrac{1}{180}\text{ hr or 20 sec}$$

A word of caution is necessary concerning the use of these equations. We stated at the beginning of this section that we were assuming that the acceleration is constant—these equations are correct *only* for the case of constant acceleration. They are an abbreviated form—a summary—of the graphical methods we have developed for these cases. But when you have a case with variable acceleration, you must apply the more powerful graphical techniques.

11–9 SUMMARY

In this chapter we have been studying straight-line kinematics—the story of motion along a straight path and of the relationships between time and position in describing this motion. After carefully defining the position of the moving object and its changes of position by the displacement, we then investigated a simple example of motion in a single direction at constant velocity. The displacement was just the product of velocity and the time of travel in this case. We could not use this simple relation when the velocity was not constant, for we did not know in advance which of the various velocities to multiply by the time. However, we were able to show how the displacement could be obtained with the help of a graph of velocity versus time, such as that shown in Fig. 11 – 33. The shaded "area" in the figure shows the displacement occurring between the times t_i and t_f.

This graph can do more for us. By drawing a tangent to the graph at any particular time, such as t, we can determine the slope of the graph at this point. This quantity is the instantaneous acceleration—the rate at which the velocity is changing. Thus, the instantaneous acceleration at the time t in Fig. 11 – 33 is

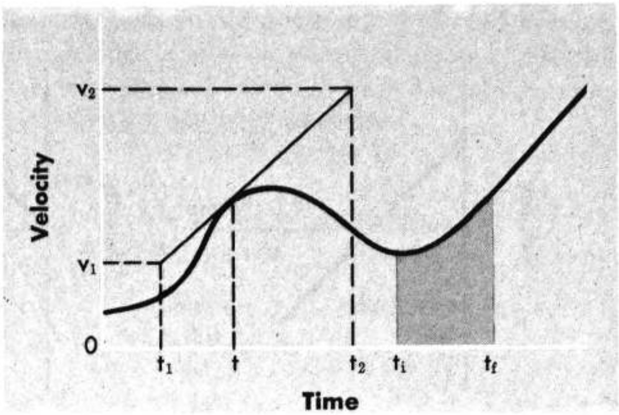
$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{v_2 - v_1}{t_2 - t_1}$$

When the curve is a straight line, as at the right of Fig. 11 – 33, the acceleration remains constant. Under such circumstances, it is very easy to compute the "area" under the graph, and we have been able to arrive at algebraic equations that tell the full story of motion at constant acceleration. The use of graphs is not limited to this simple case, however. They furnish a quite general way of studying motion.

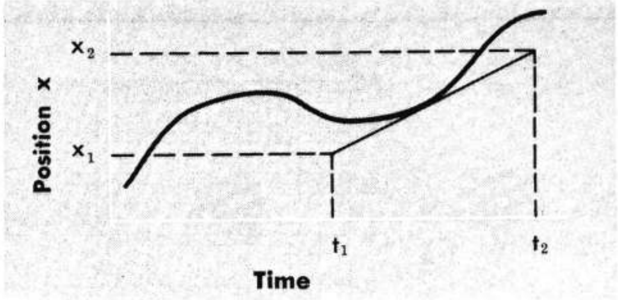
Another valuable graph is that of position versus time, such as is shown in Fig. 11 – 34. From this graph we can find the velocity as well as the displacement. The slope of this graph at any particular time is the value:

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{x_2 - x_1}{t_2 - t_1}$$

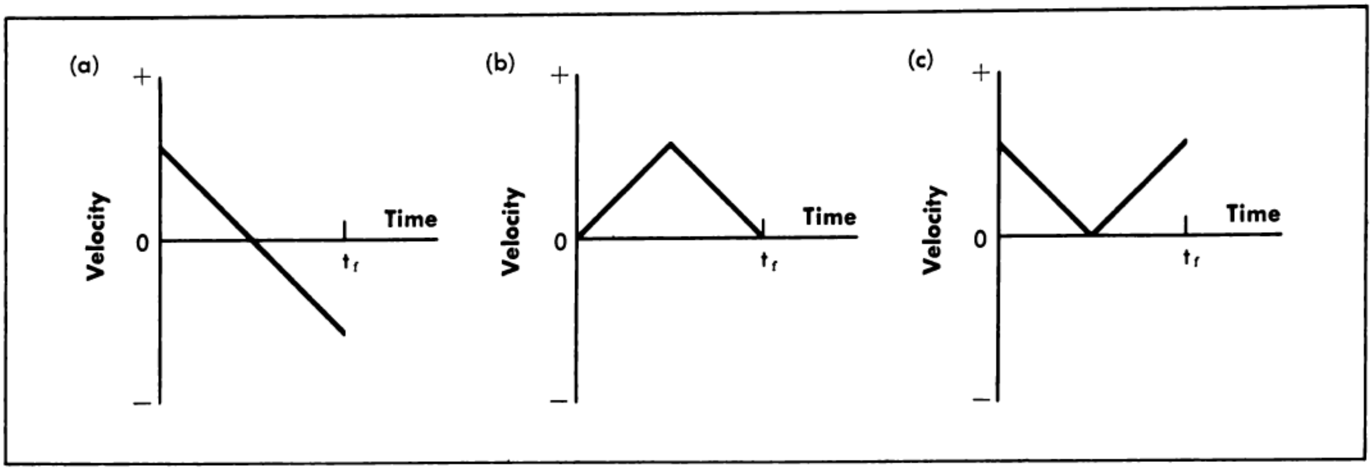
Although we have gone far in our study of motion, there are still questions we cannot answer. In the next chapter we shall get away from the restriction of motion along a straight line. We shall extend our ideas so that we can consider motion in all directions: up-down, north-south, east-west. We shall then be able to describe many of the real motions that we find in nature.



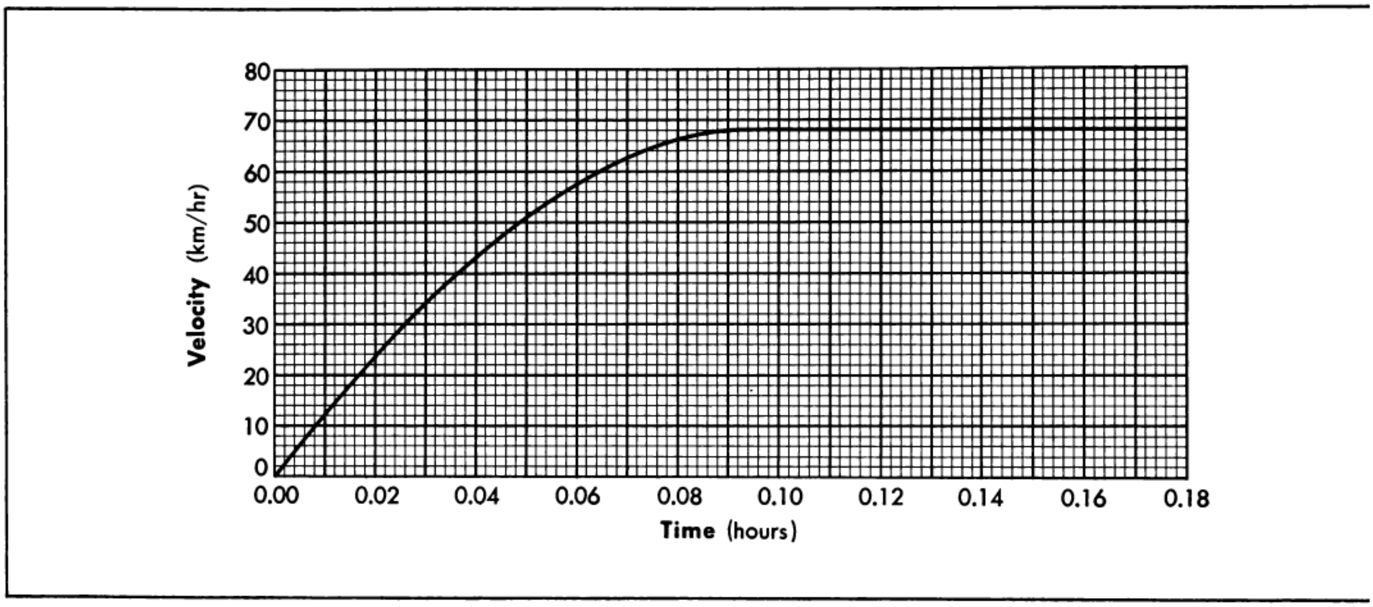
11–33 Uses of the velocity vs. time graph.



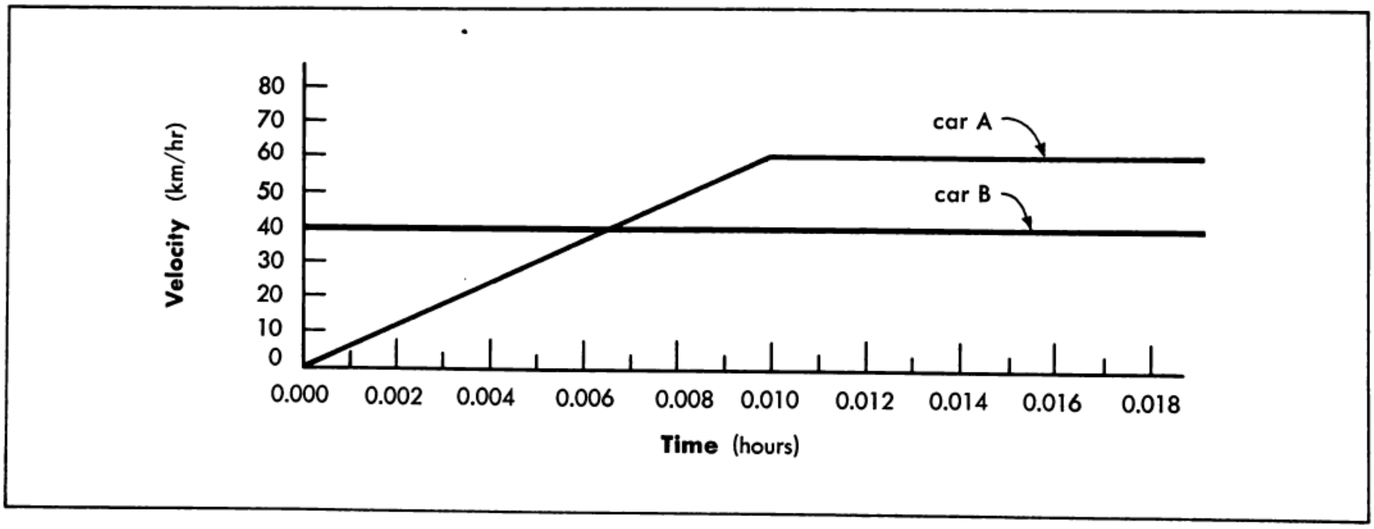
11–34 Velocity can be determined from the position-time graph.



11–35 For Question 5.



11–36 For Question 7.



11–37 For Question 8.

QUESTIONS

1.* What is the displacement of a car which travels at a steady velocity of 40 km/hr (a) for 3 hr, (b) for $\frac{1}{2}$ hr?

2. A Volkswagen and a Cadillac are driven on the same 130-km trip. The Volkswagon travels at 80 km/hr all the time. The Cadillac starts at the same time, driving at 100 km/hr, but the driver stops for ten minutes after he has traveled for half an hour.
(a) Make a graph of velocity versus time for the two cars.
(b) Which car is the first to arrive at the destination?

3. Data for a trip made in a car are given in the accompanying table:

TIME INTERVAL	DURATION OF INTERVAL *(hr)*	VELOCITY *(km/hr)*
1	0.10	20
2	0.40	60
3	0.20	20

(a) How far does the car go in the first interval?
(b) What is the displacement for the trip?
(c) Make a graph of velocity versus time for this trip.
(d) How long does it take the car to travel the first 14 km?
(e) Indicate the area on your graph which corresponds to the first 14 km of travel.

4. The following experiment was done at Marineland to determine the speed of a porpoise. A movie camera was set up to photograph through a window in the side of the tank where the porpoise passed close by just before jumping.
One sequence of pictures showed the porpoise's snout entering the left side of one frame. Eight frames later his tail was in about the same position. The camera took pictures at the rate of 24 frames/sec. The length of the porpoise is known to be 1.9 meters.
(a) What was the speed of the animal?
(b) If instead of crossing the line of sight at right angles the porpoise had crossed it at an angle of 60°, what would the calculated value of the speed be?

5.* Which graph in Fig. 11 – 35 is closest to the velocity versus time graph of a stone which is thrown straight up into the air at time $t = 0$ and returns to the earth at $t = t_f$?

6. The accompanying data show the instantaneous velocity of a car at intervals of 1 second. Plot the velocity versus time, and use your graph to answer the following questions.
(a) How fast is the car going at 2.6 sec? At 4.8 sec?
(b) How far did the car travel between the two instants in part (a)?

TIME (sec)	VELOCITY (m/sec)
0.0	10.0
1.0	12.4
2.0	14.8
3.0	17.2
4.0	19.6
5.0	22.0
6.0	24.4

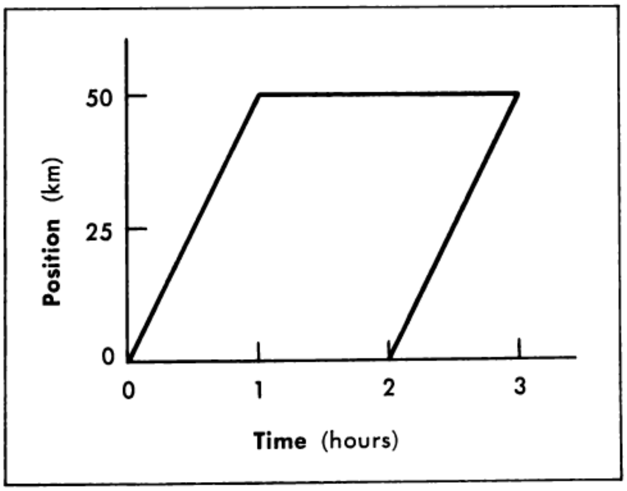
7. A train speeds up according to the velocity-time graph shown in Fig. 11 – 36. How far does it travel in the first six minutes?

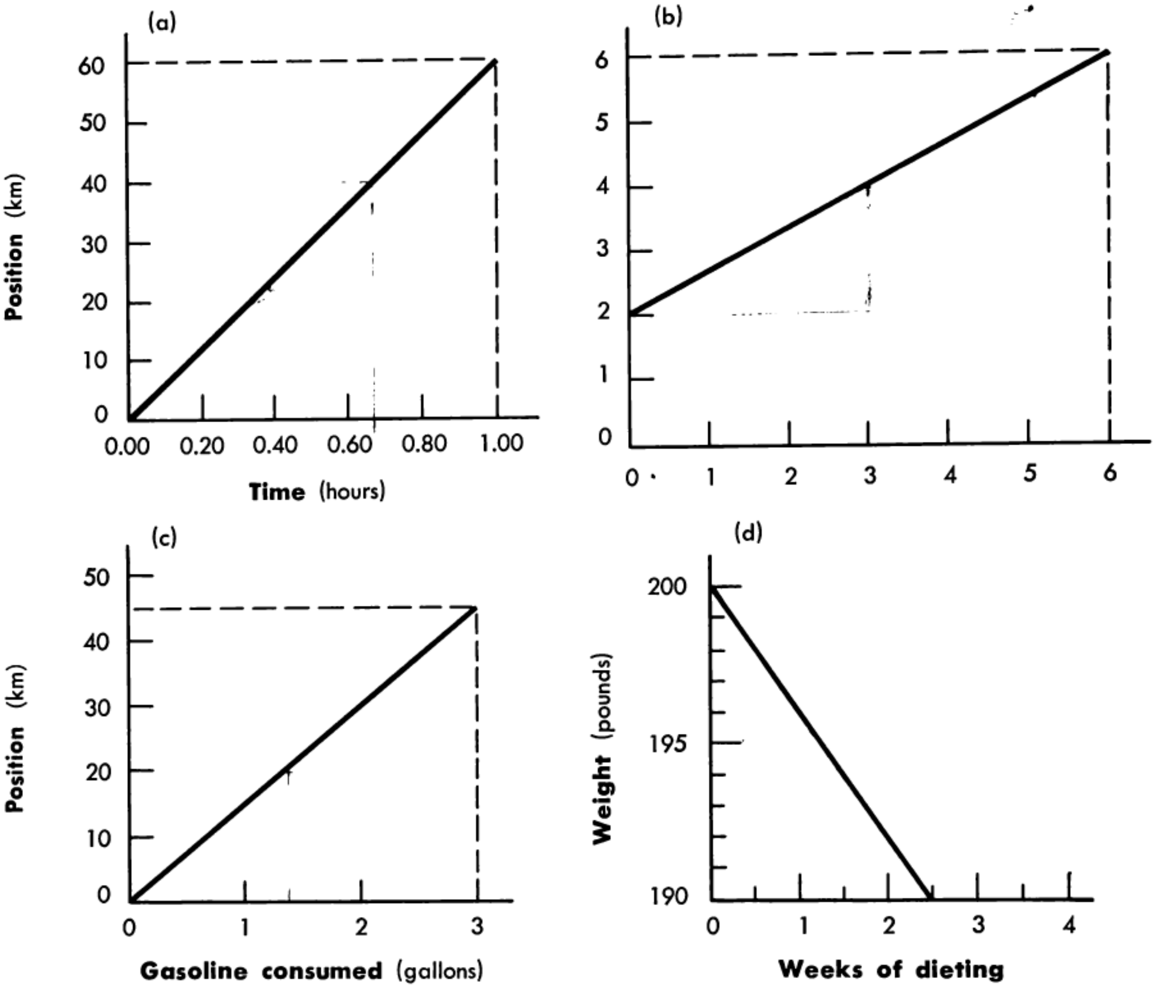
8. Car *A* is stopped at a traffic light. The light turns green and *A* starts up. Just as it does so, car *B* passes it, going at a steady velocity. Their velocity-time curves are shown in Fig. 11 – 37.
(a) How long does it take car *A* to be going as fast as car *B*?
(b) At that time, how much is car *B* ahead of car *A*?
(c) Which car is ahead, and by how much, at the end of 0.010 hour?
(d) At what time does car *A* catch up with car *B*?
(e) How far have they traveled from the traffic light by the time car *A* catches up?

9. A car is traveling along a road at a steady velocity. It passes an unmarked police car parked beside the road. The police car accelerates, overtakes the speeding car, passes it, and signals it to a stop. Sketch a graph showing the velocities of the two cars plotted against time.

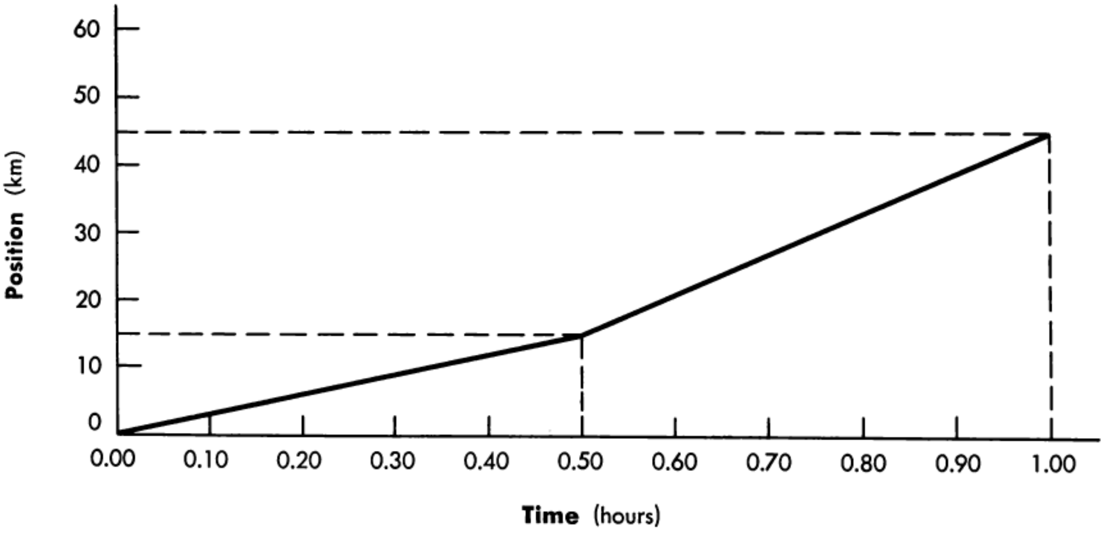
10.* Does the graph of a car trip in Fig. 11 – 38 represent a real situation? Explain.

11–38 For Question 10.





11–39 **For Question 12.**



11–40 **For Question 13.**

11. A man walks to the corner to mail a letter and comes back. Sketch graphs showing his velocity and position plotted against time.

12. Find the slopes of the graphs in Fig. 11 – 39. State the units in each case.

13. The position-time graph of a car traveling along a road is shown in Fig. 11 – 40. Make a graph of its velocity versus time.

14. A boy rode his bicycle as fast as he could from his house to a friend's house. After a short time he rode back as fast as he could. Figure 11 – 41 shows a position-time graph of his trip. Plot the velocity-time graph of the trip. From the information given and your graph, what would you give as a plausible description of the road between the two houses?

15.* What is the average velocity over the first 90 sec in the graph of Fig. 11 – 25?

16. The position-time graph of a car is shown in Fig. 11 – 42.

(a) At what time is the car going at the greatest velocity?

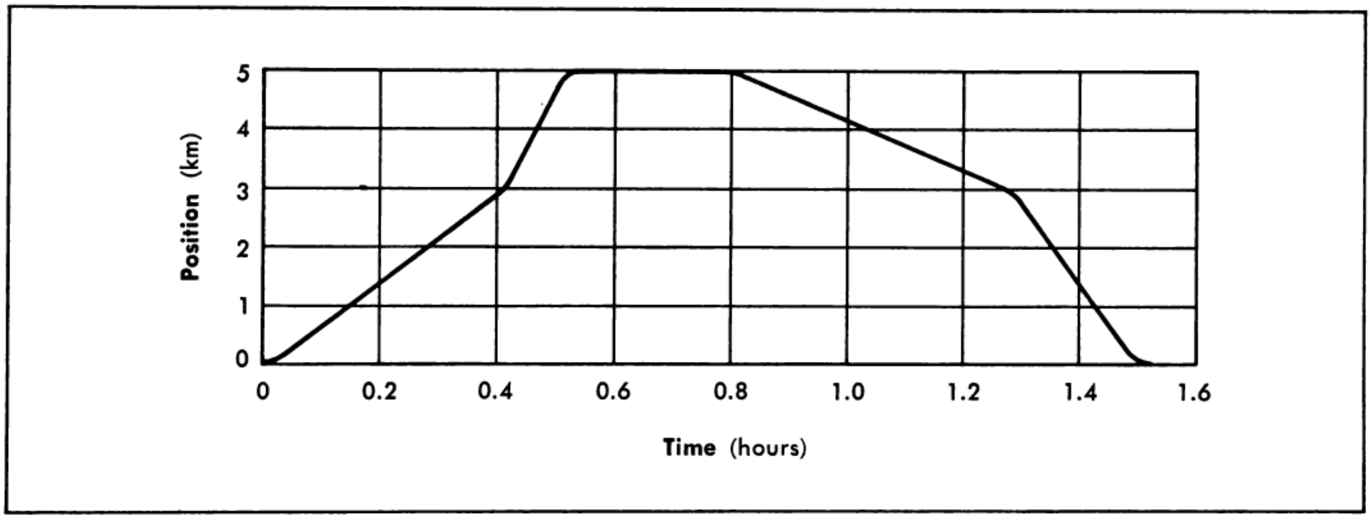
(b) How fast is it traveling at that time?

(c) How fast is the car going at 0.70 hour?

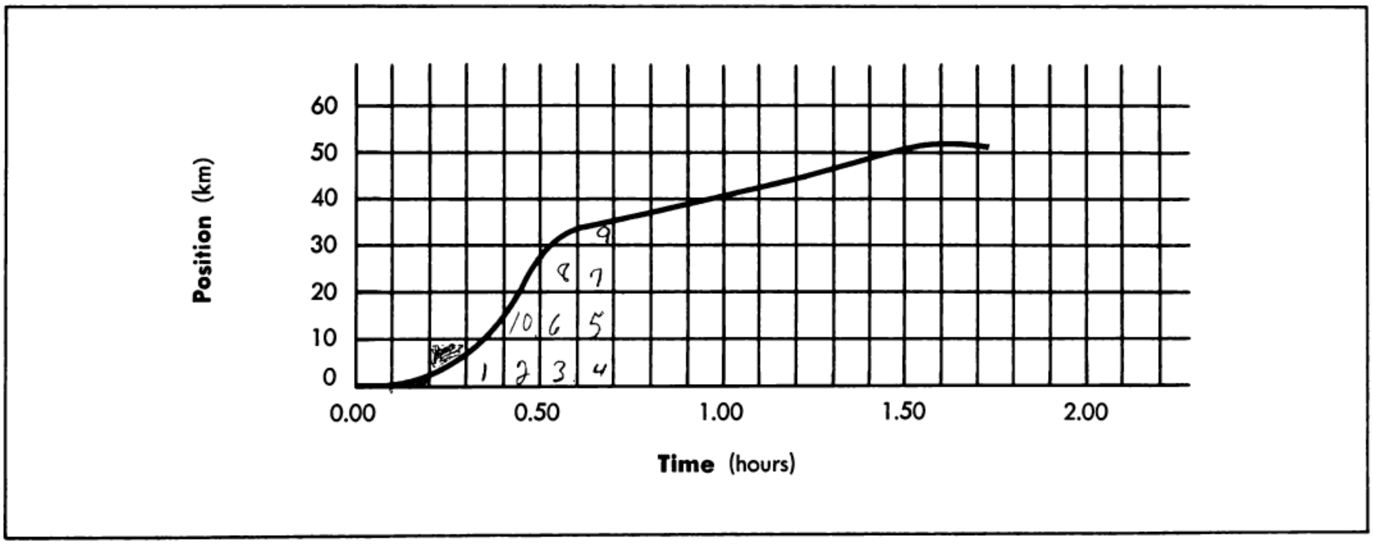
(d) What is the average velocity for the first 0.70 hour?

17. A train travels 100 km/hr for 0.52 hr, 50 km/hr for the next 0.24 hr, and then 110 km/hr for the next 0.71 hr. What is its average velocity for the trip?

18.* In Fig. 11 – 30, is the velocity zero at any time on the graph? Is the acceleration zero at any time on the graph?



11–41 For Question 14.

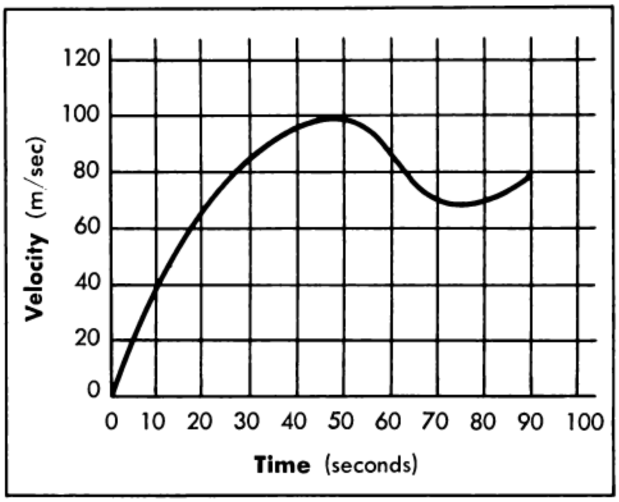


11–42 For Question 16.

19. A car driving along a highway slows down as it enters a small village. It is stopped by a traffic light in the center, goes on to the edge of the village, and speeds up again. Sketch graphs, one above the other, showing its velocity, its position, and its acceleration plotted against time.

20. A car going at 20 km/hr accelerates to 60 km/hr in 6.0 seconds. What is the average acceleration?

21. From the graph of velocity versus time for a car shown in Fig. 11 – 43, deduce the graph of acceleration versus time.



11–43 For Question 21.

22. Make a graph of position versus time for the car of Question 21.

23.* In what portion of the graph in Fig. 11 – 33 can you use the equations of Section 11 – 8?

24. A bobsled has a constant acceleration of 2.0 m/sec² starting from rest.
    (a) How fast is it going after 5.0 seconds?
    (b) How far has it traveled in 5.0 seconds?
    (c) What is its average velocity in the first 5.0 seconds?
    (d) How far has it traveled by the time its velocity has reached 40 m/sec?

25. A car, initially traveling at uniform velocity, accelerates at the rate of 1.0 m/sec² for a period of 12 seconds. If the car traveled 190 meters during this 12-second period, what was the velocity of the car when it started to accelerate?

26. Assume that the express train mentioned at the very beginning of the chapter stops with uniform acceleration.
    (a) How long does it take to stop?
    (b) What is the acceleration during the braking?
    (c) Plot the position of each train versus time on the same graph and from this find whether or not a collision takes place.

27. A pedestrian is running at his maximum speed of 6.0 m/sec to catch a bus stopped by a traffic light. When he is 25 meters from the bus the light changes and the bus accelerates uniformly at 1.0 m/sec². By use of a graph find either (a) how far he has to run to catch the bus or (b) his frustration distance (closest approach).

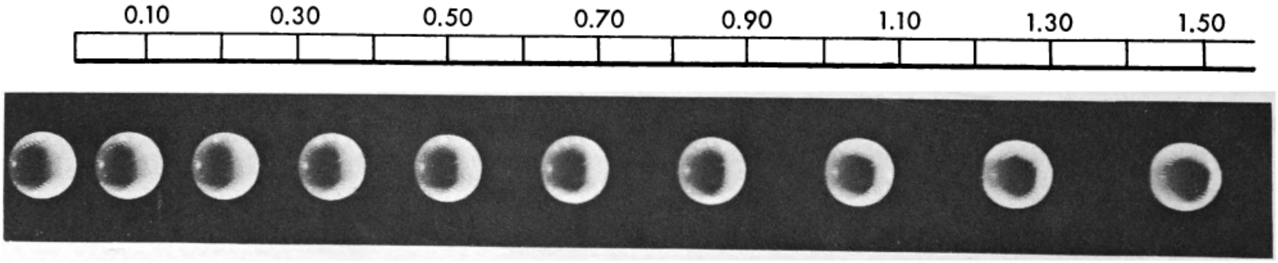
28. Figure 11 – 44 is a multiple-flash photograph of a moving ball taken at 1/30 sec intervals. The ball is moving from left to right, and the zero point on the scale lines up with the right-hand edge of the ball's initial position.
    (a) Plot a graph of position against time to describe this motion.
    (b) From your graph in (a), construct a velocity-time graph.
    (c) What does your velocity-time graph tell you about the acceleration of the moving ball?

29. A car accelerates from rest at a constant rate of 3 m/sec² for 10 seconds. On rectangular graph paper, draw a graph of (a) acceleration as a function of time, (b) velocity as a function of time, (c) displacement as a function of time.

30. Repeat Question 29, but plot the three graphs on log-log paper.

31. A rocket which placed a satellite in orbit attained a velocity of 2.90 × 10⁴ km/hr in 2.05 minutes.
    (a) What was the average acceleration in km/hr-sec? In m/sec²?
    (b) If the rocket had enough fuel to maintain this same acceleration for an hour, what velocity would it have at the end of the hour if it started from rest?
    (c) How far would it travel during this hour?

11–44 For Question 28.



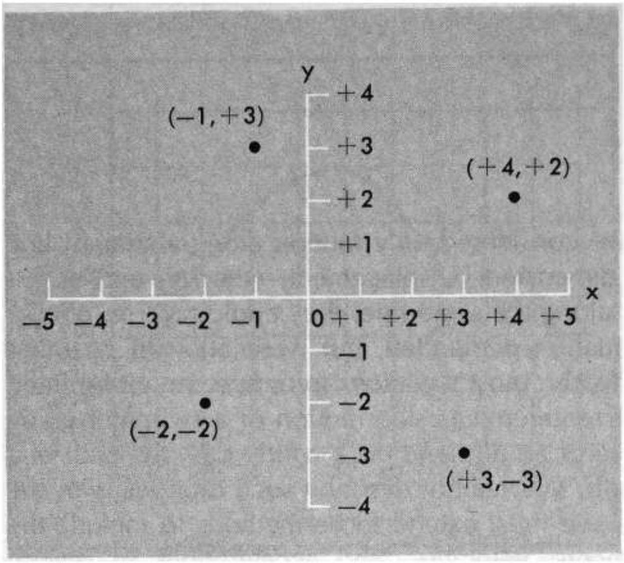
12 Motion in Space

IN THE LAST chapter we considered only motion along a straight line —motion in only one dimension. Displacement, velocity, and acceleration all had to point along the same line; they could take on plus or minus values as the situation demanded, but were allowed no other freedom of direction. Clearly, most situations in nature cannot be fitted into such a strict set of requirements: the motion of a swimmer going across a river, the course of an airplane on a windy day, the path of a satellite orbiting the earth. We cannot describe such motions with the methods of Chapter 11. We must extend those methods to include the possibility of displacements, velocities, and accelerations in several different directions, perhaps all occurring at the same time; we must learn to treat motion in a three-dimensional world. However, much of the extension required to go from one to three dimensions is already present in a description of two-dimensional motion. We shall start, then, with motion on a plane surface.

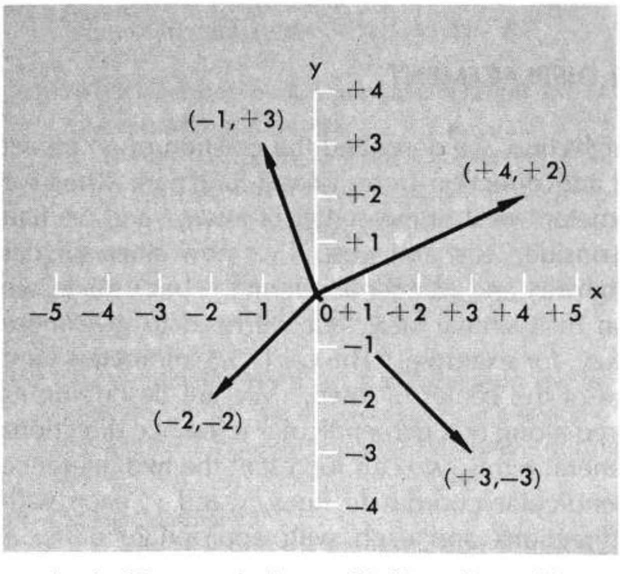
12–1 POSITION AND DISPLACEMENT

For motion along a straight line, we described the position of an object in terms of its distance and direction from a fixed landmark. Thus we said, "The car is 5 kilometers west of the center of town," and we had only two directions to consider, east and west. If we now allow the car to turn off the straight highway and wander about on the back roads, we can describe its position in a similar way, but we need to give more information. We might say, for example, "The car is 5.5 kilometers west and 1.2 kilometers north of the center of town." We are describing its *two* coordinates, measured along two perpendicular reference directions in the plane. In more general terms, we can represent the two reference directions by two perpendicular coordinate lines, x and y, each with positive and negative directions and each with appropriate units. A position on the plane is then given by two coordinate numbers; the x coordinate is stated first, the y coordinate second. Several points are located by their coordinates in Fig. 12 – 1 (a).

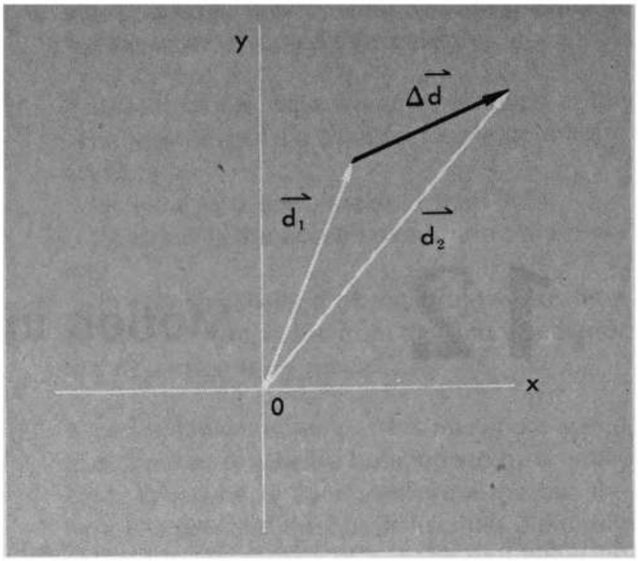
Another method for locating a point on the plane is by using a directed line segment starting at the origin of the coordinates and ending at the point [Fig. 12 – 1 (b)]. The length of the segment, in appropriate units, is just the straight-line distance between the origin and the point; the direction of the segment is the direction of the point from the origin. The directed line segment, representing both the magnitude and the direction in space of some physical quantity, is called a *vector* and will be symbolized by a boldface letter with a half-arrow above it. The directed segment which describes the position of an object relative to an origin is the *position vector*, $\vec{\mathbf{d}}$.



12–1a Position on a plane can be specified by the two coordinate numbers of the point: first, the distance out from the origin along the x axis; second, the distance up, parallel to the y axis.



12–1b Position can also be specified by a directed line segment, or position vector, from the origin to the point.



12–2 The change of position from the point specified by $\vec{d}_1$ to the point specified by $\vec{d}_2$ can be represented by the displacement vector $\Delta\vec{d}$, directed from the initial point to the final point.

To discuss *motion* on a plane, we need to be able to describe the change of position of an object as time progresses. In Fig. 12 – 2, if we wish to describe a motion from position $\vec{\mathbf{d}}_1$ to position $\vec{\mathbf{d}}_2$, for example, we see that we can again use a directed line segment, or vector This vector starts at the point described by $\vec{\mathbf{d}}_1$ and goes to the point described by $\vec{\mathbf{d}}_2$. But "change of position" is what we called "displacement" in Chapter 11. We have here, then, the *displacement vector*, $\Delta\vec{\mathbf{d}}$, describing change of position on a plane. Now we have the means for extending all our definitions of the previous chapter to motions in a plane. But first we must investigate the properties of vectors and find out how to work with them.

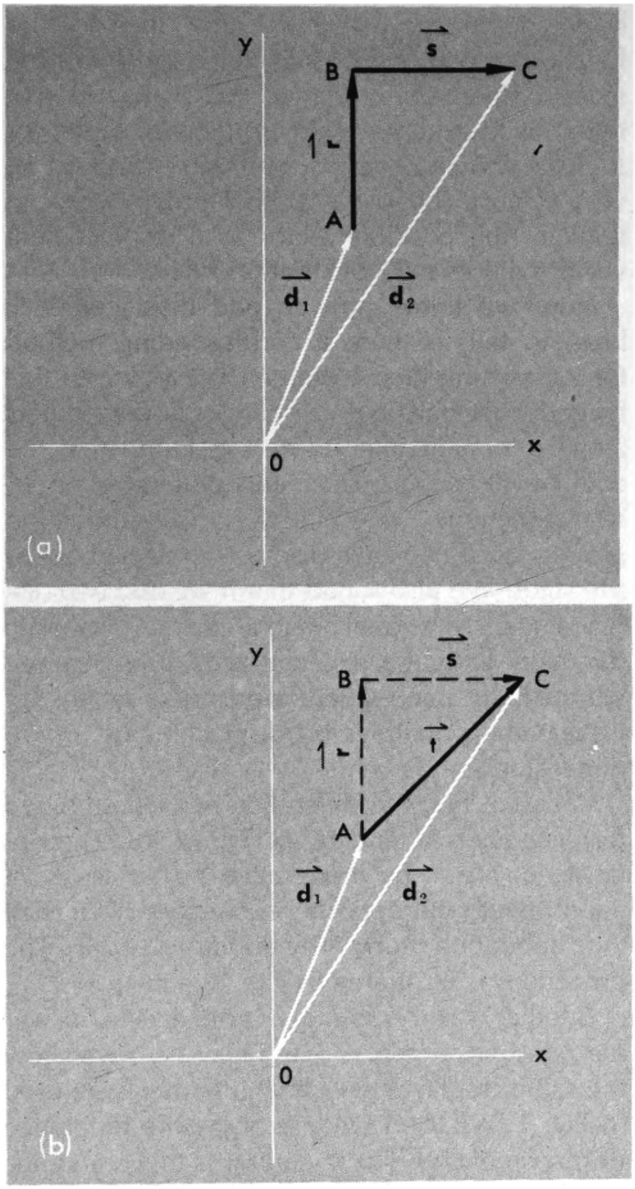
12–2 ADDITION AND SUBTRACTION OF DISPLACEMENTS

Displacements along a line were given by positive and negative numbers; when such displacements were added, they simply obeyed the rules of ordinary arithmetic. In adding displacements on a plane we meet a new problem. In Fig. 12 – 3 (a) we start at a point *A*, described by the position vector $\vec{\mathbf{d}}_1$, make a 5.0 cm displacement in the $+y$ direction to the point *B*, and then follow that with a 5.0 cm displacement in the $+x$ direction to the point *C*. A displacement directly from *A* to *C* gives the same result as the two displacements separately [Fig. 12 – 3 (b)]. Therefore, we can write

$$\text{displacement } AB + \text{displacement } BC = \text{displacement } AC$$

or, using vector notation and calling the displacements $\vec{\mathbf{r}}$, $\vec{\mathbf{s}}$, and $\vec{\mathbf{t}}$ respectively, we have

$$\vec{\mathbf{r}} + \vec{\mathbf{s}} = \vec{\mathbf{t}}$$

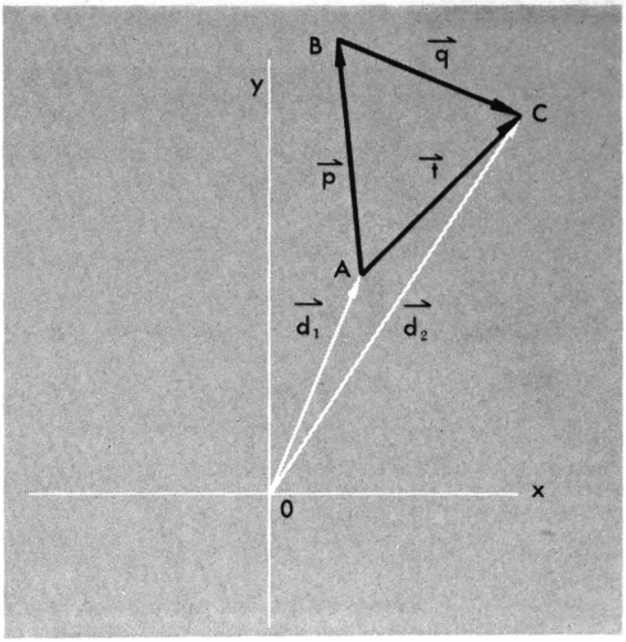


12–3 The displacement $\vec{r}$ from point A to point B can be added to the displacement $\vec{s}$ from point B to point C by the use of scale drawings. The resulting displacement is the same as a single displacement $\vec{t}$ from A to C: $\vec{r} + \vec{s} = \vec{t}$.

Of course, the individual displacements need not be at right angles; they can have any directions, as in Fig. 12 – 4. No matter what directions they have, we can always add them by this construction. We therefore see that the rule for adding displacements depends on geometry in the two-dimensional case, rather than on arithmetic, as was the case in one dimension.

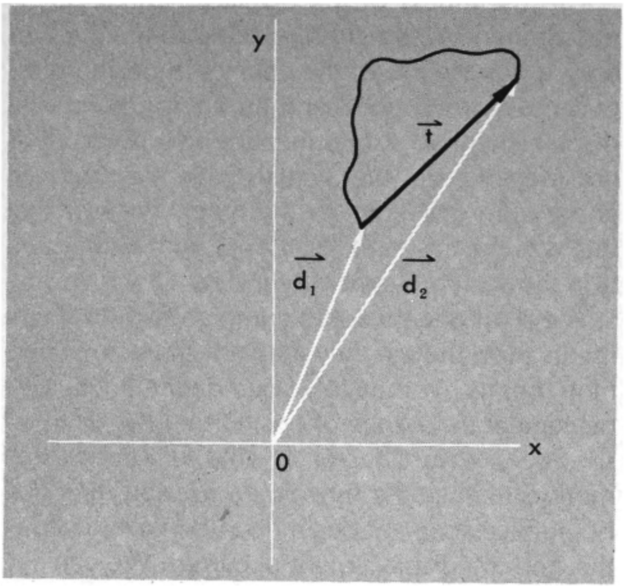
The usefulness of the displacement vector is not limited to straight-line trips. Since it gives only a description of the change in position between a starting point and an ending point, the displacement vector connecting two points does not depend on the actual path traversed in going from one point to the other. We saw that this was the case in Figs. 12 – 3 and 12 – 4, and it is shown more generally in Fig. 12 – 5.

We shall often want to compare two displacements even though they may not share any common points. Fortunately, our definition of displacement as *change* of position allows us to do so. We saw in Chapter 11 that for straight-line motion neither the magnitude nor the direction of the displacement depended upon where along the line the displacement occurred. We can see that the displacement is independent of position

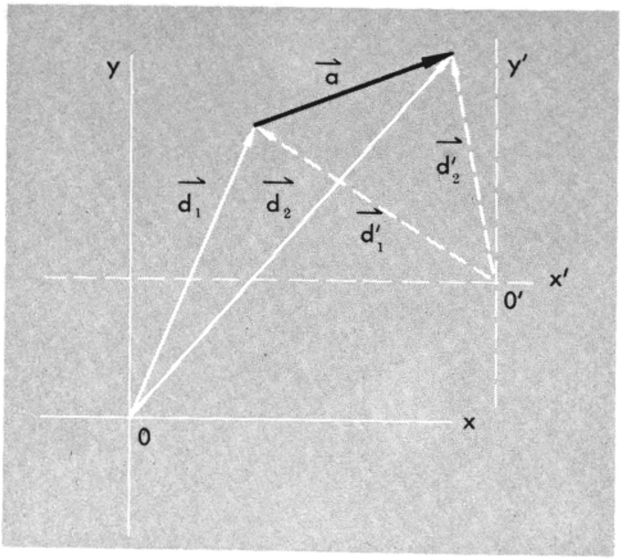


12–4 Displacements need not be at right angles to be added vectorially. Here $\vec{p} + \vec{q} = \vec{t}$ also.

for motion on a plane also. In Fig. 12 – 6, the end points of the displacement $\vec{\mathbf{a}}$ are designated by two sets of position vectors: $\vec{\mathbf{d}}_1$ and $\vec{\mathbf{d}}_2$ measured from the origin *O*, and $\vec{\mathbf{d}}_1'$ and $\vec{\mathbf{d}}_2'$ measured from the origin *O′*. Clearly, the displacement $\vec{\mathbf{a}}$ stands quite independent of the choice of origin, and hence of the actual position vectors; it depends only on the *change* of position. As an example, when you displace a book one meter east and one meter north, it ends up $\sqrt{2}$ meters northeast of its starting point, no matter where that starting point was, and no matter what point you chose as origin from which to make your measurements.



12–5 The displacement between two points does not depend on the actual path traversed in going from one point to another.



12–6 The displacement between two points does not depend on the choice of origin from which the position vectors are measured.

We can therefore define displacement without specifying where either end is, and we shall add displacements by placing them "head to tail" as in Fig. 12 – 4. Thus, if we want to find the sum of the two vectors shown in Fig. 12 – 7 (a), we can move one of them "parallel to itself" until its "tail" coincides with the "head" of the other, as shown in Fig. 12 – 7 (b). This means that two displacements are the same if their magnitudes (lengths) and directions are the same, no matter where we draw them. If the location of one end of a vector should turn out to be important sometime, we shall add that information when we use it. (For example, for a given choice of origin, one end of a position vector must always be at that origin. We can think of the position vector as a displacement vector with its tail stuck at the origin.)

Now we know how to add displacements, head to tail; is there a corresponding method for subtracting them? We can get a clue to this by going back again to the case of displacements along a straight line. There we found that the addition of two displacements pointed in opposite directions was effectively a subtraction process since two such displacements had opposite signs. We also found that if we add two displacements of equal magnitude by opposite direction, we get a resultant zero displacement; we are right back where we started. Let us try these same principles on displacements in two dimensions.

We shall find the difference between the two displacements $\vec{\mathbf{a}}$ and $\vec{\mathbf{b}}$ in Fig. 12 – 8. The opposite of $\vec{\mathbf{b}}$ is $-\vec{\mathbf{b}}$, represented by an arrow of equal length but opposite in direction. We shall then subtract $\vec{\mathbf{b}}$ from $\vec{\mathbf{a}}$ by adding $-\vec{\mathbf{b}}$ to $\vec{\mathbf{a}}$. The procedure is illustrated in Fig. 12 – 8.

Another way to subtract $\vec{\mathbf{b}}$ from $\vec{\mathbf{a}}$ is to answer the question $\vec{\mathbf{a}} = \vec{\mathbf{b}} + ?$. In Fig. 12 – 9 we show the same displacements $\vec{\mathbf{a}}$ and $\vec{\mathbf{b}}$ that were used in Fig. 12 – 8, and we also show the difference displacement $\vec{\mathbf{a}} - \vec{\mathbf{b}}$. Clearly it is this difference displacement that must be added to $\vec{\mathbf{b}}$ to produce $\vec{\mathbf{a}}$. Notice that the displacement $\vec{\mathbf{a}} - \vec{\mathbf{b}}$ in Fig. 12 – 9 is parallel to the corresponding displacement of the same length and direction in

Fig. 12 – 8. Therefore, they are the same. The two methods give the same answer; use the method that seems most convenient for you.

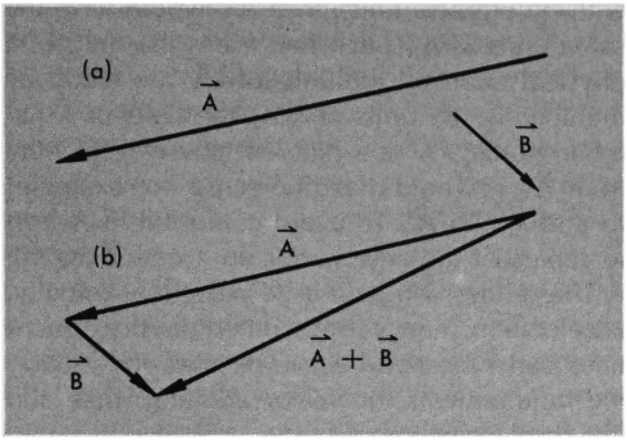
We can use the second method to show that the displacement vector $\Delta\vec{\mathbf{d}}$ is really just the difference between the final and the initial position vectors, as we expect from its name: change of position. We see in Fig. 12 – 2, for example, that the displacement vector $\Delta\vec{\mathbf{d}}$ can be considered as being added to the position vector $\vec{\mathbf{d}}_1$ to produce the position vector $\vec{\mathbf{d}}_2$:

$$\vec{\mathbf{d}}_1 + \Delta\vec{\mathbf{d}} = \vec{\mathbf{d}}_2$$

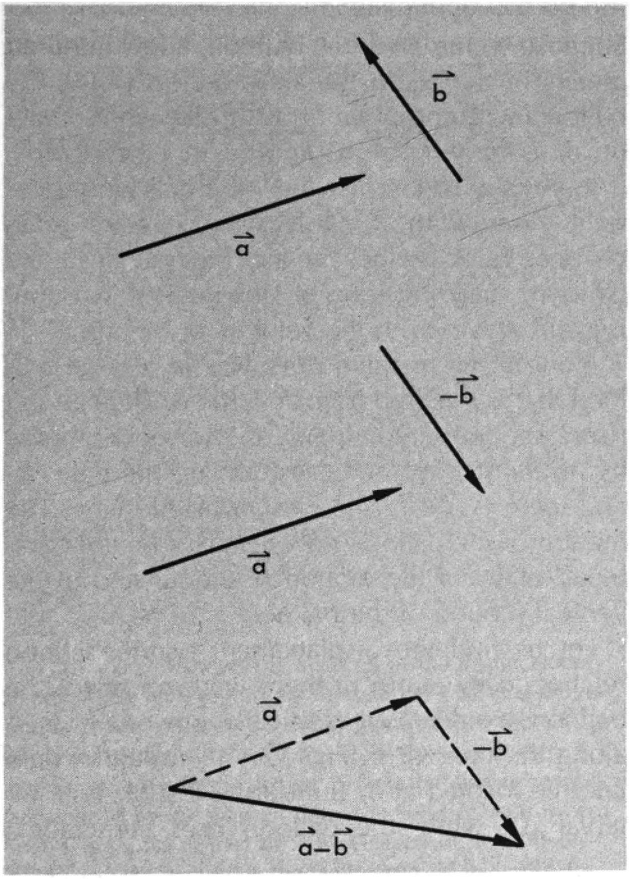
Hence we can write

$$\Delta\vec{\mathbf{d}} = \vec{\mathbf{d}}_2 - \vec{\mathbf{d}}_1$$

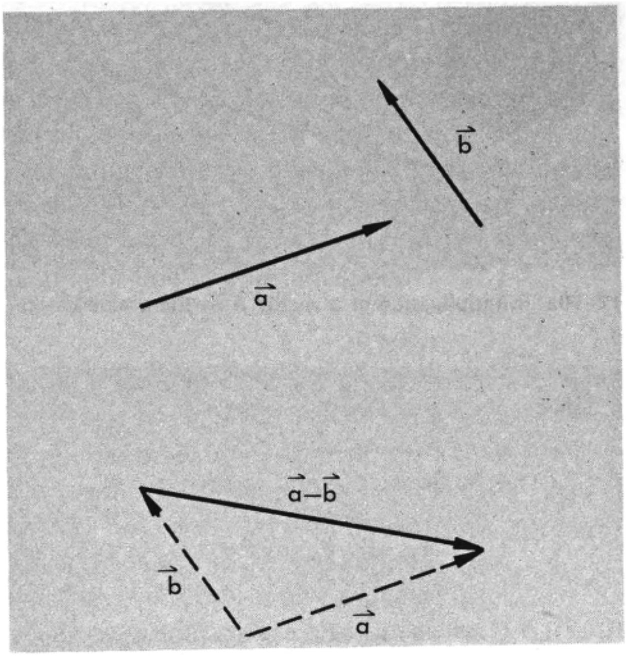
We have used displacements and the addition of displacements as examples of the new quantitites we call vectors. Several quantities of importance in our study of physics will also prove to be vectors. Vectors are any quantities which add and subtract like displacements. We shall always represent vectors by directed line segments having lengths proportional to the magnitudes and having directions indicated by arrowheads. Notice that vectors are quantities which cannot be described by single numbers. Scalars, on the other hand, are physical quantities which can be specified by single numbers (and the appropriate physical units). The temperature of a room is a good example of a scalar physical quantity, just as a displacement of 3 kilometers northeast is a good example of a vector quantity.



12–7 Addition of two vectors after moving one parallel to itself.



12–8 To find the difference of two vectors we add the negative of the second to the first.



12–9 We can also find $\vec{a}$-$\vec{b}$ by drawing $\vec{a}$ and $\vec{b}$ tail to tail, then drawing a vector from the head of $\vec{b}$ to the head of $\vec{a}$. You can check that you have drawn this vector in the right direction by remembering that $\vec{b} + (\vec{a}\text{-}\vec{b}) = \vec{a}$.

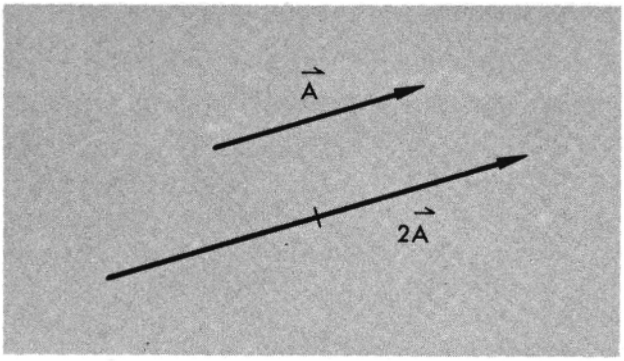
12–3 MULTIPLYING VECTORS BY NUMBERS AND SCALARS

Suppose that you are given any vector $\vec{\mathbf{A}}$. What do you think should be meant by $2\vec{\mathbf{A}}$? Because multiplying a quantity by two means adding the quantity to itself, it is logical to define

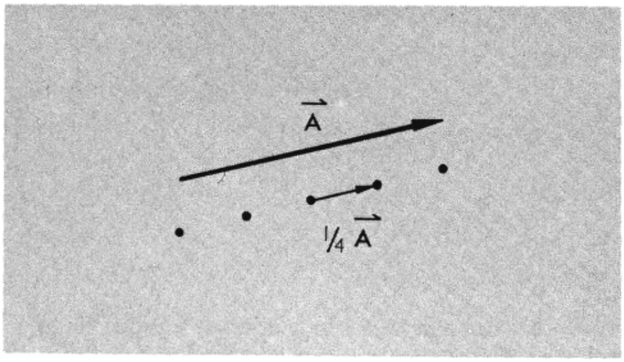
$$2\vec{\mathbf{A}} = \vec{\mathbf{A}} + \vec{\mathbf{A}}$$

that is, $2\vec{\mathbf{A}}$ is a vector twice as long as $\vec{\mathbf{A}}$ pointing in the same direction [Fig. 12 – 10 (a)]. Similarly, just as $\frac{1}{4}x$ means a number which added four times over gives x, we define $\frac{1}{4}\vec{\mathbf{A}}$ as a vector which added four times over gives $\vec{\mathbf{A}}$; in other words, $\frac{1}{4}\vec{\mathbf{A}}$ is a vector in the same direction as $\vec{\mathbf{A}}$, but only $\frac{1}{4}$ as long [Fig. 12 – 10 (b)]. Notice that multiplication of $\vec{\mathbf{A}}$ by $\frac{1}{4}$ is the same as dividing $\vec{\mathbf{A}}$ by 4, so we can expect division by a number to follow the same rules as multiplication by a number. In general, $k\vec{\mathbf{A}}$, where k is any positive number, means a vector parallel to $\vec{\mathbf{A}}$ and k times as long. What would you expect if k is negative?

Briefly then, multiplying or dividing a vector by an ordinary number means multiplying or dividing the magnitude by the absolute value* of that number, while leaving the direction unchanged if the number is positive, or reversing the direction if the number is negative.



12–10a Multiplication of a vector $\vec{A}$ by the scalar $k = 2$.



12–10b Multiplication of a vector $\vec{A}$ by the scalar $k = 1/4$.

Now we extend this rule to cases in which both the scalar k and the vector $\vec{\mathbf{A}}$ have physical units, such as seconds, kilometers, 1/second, kilometers/second, etc. If k were a pure number, with no physical units, then $k\vec{\mathbf{A}}$ would have the same units as $\vec{\mathbf{A}}$. But when k has the units of a physical quantity, the units of $k\vec{\mathbf{A}}$ are found by multiplying the units of k by the units of $\vec{\mathbf{A}}$. In other words, $k\vec{\mathbf{A}}$ is a different physical quantity from $\vec{\mathbf{A}}$; it would make no sense, for example, to add $k\vec{\mathbf{A}}$ to $\vec{\mathbf{A}}$. To avoid confusion, it is best to represent the new vector on a new diagram.

These rules will help us to prove that velocity, acceleration, and various other physical quantities can be called vectors because they satisfy the requirements for vector addition: they add like displacements.

12–4 VELOCITY VECTORS IN SPACE

Suppose we are in a free balloon, a few hundred meters in the air. If the air is perfectly still, the balloon will not move at all. If the whole mass of air is moving at 10 km/hr—in other words, if there is a 10 km/hr wind—the balloon moves right along with it. Only when the air mass changes velocity does the air "blow against" the balloon, speeding it up or slowing it down until it again moves with the velocity of the air.

Now let us imagine ourselves in a large balloon that is released from an airfield. The balloon rises at a constant velocity of 5 m/sec as shown by a rate-of-climb indicator aboard the balloon, and there is a 3 m/sec west wind blowing. The motion of the balloon now results from the combined effect of the motion of the air and of the vertical motion of the balloon.

Let us find the displacement of the balloon relative to the earth. If there were no wind, the balloon would move vertically upward from A along the line AB in Fig. 12 – 11 (a) and would later be located at B. But during a short time interval Δt, while the balloon rises vertically a

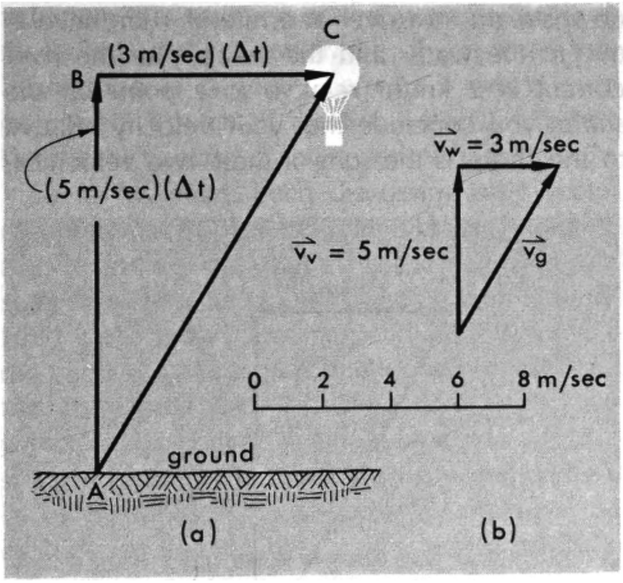
---

*The *absolute value* of a number is the number without its algebraic sign. Thus, the absolute value is always positive. For example, the absolute value of −3.5 is 3.5.

distance (5 m/sec) Δt from A to B, the air has moved the balloon east a distance (3 m/sec) Δt and so the balloon arrives at the position C.

The addition of the two displacement vectors $\overrightarrow{AB}$ and $\overrightarrow{BC}$ leads to the vector $\overrightarrow{AC}$, the displacement of the balloon relative to the ground. These three vectors form a triangle. On dividing each of these displacements by the elapsed time, Δt, we get a similar triangle representing the corresponding velocities (since vectors divided by scalars are still quantities in the same direction, even though these may have different lengths and different units). We now see that these velocities add exactly as the displacements, so we have proved that velocity is a vector.

In Fig. 12 – 11 (b) the velocity diagram is shown. The vertical velocity $\vec{\mathbf{v}}_v$ of the balloon relative to the ground is plotted in the same direction as $\overrightarrow{AB}$. It is a vertical line of length proportional to 5 m/sec. To this velocity we must add the velocity of the wind relative to the earth $\vec{\mathbf{v}}_w = 3$ m/sec pointed east). We then obtain the velocity $\vec{\mathbf{v}}_g$ of the balloon relative to the ground. Both its direction and its magnitude can be obtained directly from its graphical representation.



12–11 **Part (a) shows the addition of the displacement vector $\overrightarrow{AB}$, which represents the balloon's vertical motion, and $\overrightarrow{BC}$, which represents the displacement of the air. Their sum is the vector $\overrightarrow{AC}$, the displacement of the balloon relative to the ground. When the displacements are divided by the elapsed time, the velocity vectors are found as shown in part (b). The velocity of the balloon relative to the ground is the vector sum of its vertical velocity relative to the ground and the velocity of the wind relative to the ground.**

Because $\vec{\mathbf{v}}_v$ and $\vec{\mathbf{v}}_w$ are at right angles in this example, it is easy to calculate the magnitude of the velocity. Using the Pythagorean relationship, we see that

$$v = \sqrt{v_v^2 + v_w^2}$$
$$v = \sqrt{(5 \text{ m/sec})^2 + (3 \text{ m/sec})^2}$$
$$= 5.8 \text{ m/sec}$$

Notice that the various *speeds* are indicated by v, v_v, and v_w, the same symbols that are used for the velocities, except that they do not carry the vector sign over them. This is a very convenient and commonly used notation.

Since we have shown that velocity is a vector quantity, we should therefore write a general definition of velocity in vector notation. For a displacement $\Delta\vec{\mathbf{d}}$ occurring in a time interval Δt, the average velocity during the interval is

$$\vec{\mathbf{v}}_{\text{ave}} = \frac{\Delta\vec{\mathbf{d}}}{\Delta t}$$

Clearly $\vec{\mathbf{v}}_{\text{ave}}$ has the direction of $\Delta\vec{\mathbf{d}}$.

If the time interval is taken ever smaller, in the limit the instantaneous velocity is

$$\vec{\mathbf{v}} = \lim_{\Delta t \to 0} \frac{\Delta\vec{\mathbf{d}}}{\Delta t}$$

The direction of $\vec{\mathbf{v}}$ is the direction of $\Delta\vec{\mathbf{d}}$ for a very small displacement; it is therefore along, or tangent to, the path.

In the example of the balloon we assumed constant velocities for the vertical motion of the balloon and for the wind. In order for an instantaneous velocity to be called "constant," two things must be constant: its magnitude *and* its direction. A body moving with constant velocity would, therefore, move in one direction along a straight line at constant speed.

12–5 COMPONENTS OF VECTORS

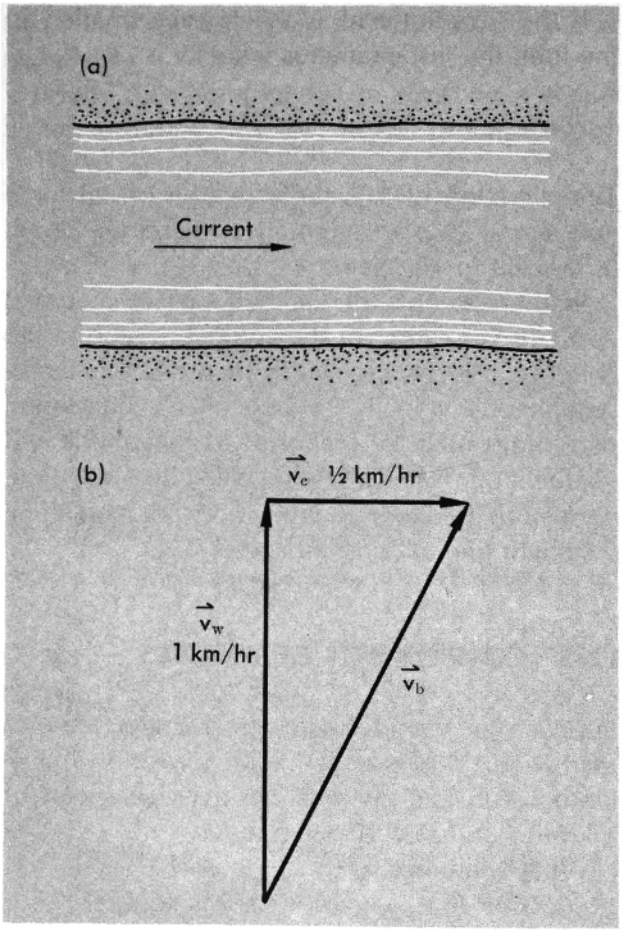
Suppose you want to swim across a river $\frac{1}{4}$ kilometer wide. The current in the river flows at $\frac{1}{2}$ km/hr. You can swim at an average speed of 1 km/hr, but only for 15 minutes.

What is the best plan to follow?

Suppose that you swim so as to keep your body always pointed at right angles to the bank. You will then be moving through the water perpendicular to the current, which will carry you steadily downstream. What will your velocity be with respect to the bank of the river? If we represent your velocity through the water perpendicular to the bank by $\vec{v}_w$ and that of the current by $\vec{v}_c$, your velocity with respect to the bank is $\vec{f}_b$ in Fig. 12 – 12.

You can see that the magnitude of $\vec{v}_b$ is greater than that of $\vec{v}_w$, but does this get you across the river faster? No, your motion perpendicular to the banks of the river (which determines the time of crossing) is still the same. You will just make it. The velocity $\vec{v}_b$ tells us only that you will reach the other side downstream from the starting point.

The velocity $\vec{v}_b$ is made up of two parts, perpendicular to each other; these are called the *rectangular vector components* of $\vec{v}_b$. In this case, only the vector component $\vec{v}_w$ perpendicular to the bank affects the time of crossing. The other component $\vec{v}_c$ does not affect your progress to the other shore. Such rectangular components are mutually independent; they do not affect each other. No matter what the $\vec{v}_c$ is, you would still make it in 15 min.



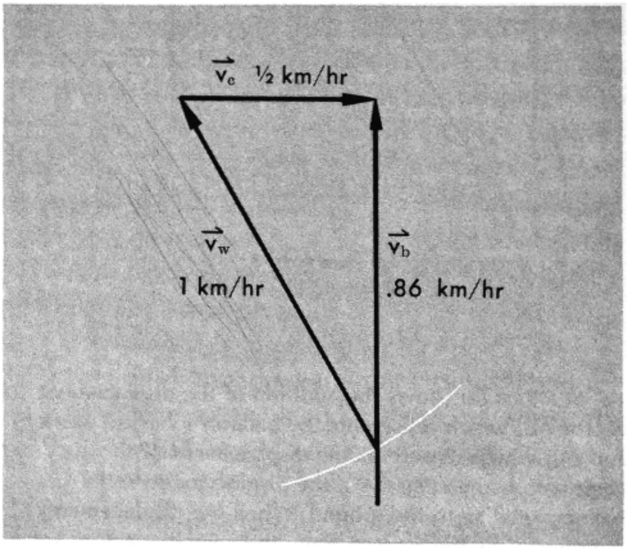
12–12 If you swim across a river with a velocity $\vec{v}_w$ through the water at right angles to the current $\vec{v}_c$, your velocity relative to the bank will be $\vec{v}_b$. Does the current help you reach the other side?

Suppose you are faced with a tougher problem: you can get ashore only at a beach which is directly opposite your starting point. Then your velocity $\vec{v}_b$ relative to the bank must be at right angles to it. This says that the rectangular vector component of $\vec{v}_b$ parallel to the bank must be zero. You must point your body upstream so that the rectangular vector component of $\vec{v}_w$ (Fig. 12 – 13) parallel to the bank is equal in magnitude and opposite in direction to $\vec{v}_c$.

Now your velocity across the river is only about 0.86 km/hr and you won't make it. Perhaps you can wade the rest of the way.

We have seen that heading upstream makes it harder to get across the river. What would happen if you headed somewhat downstream? Of course, you won't get to the beach, but will you even get across the river? Do you agree that a current across your intended path cannot possibly help you?

In adding velocities we make an important assumption which in some cases can lead to errors. For example, if you are rowing a boat downstream in a river at a rate of 1 km/hr *relative to the water* and the velocity of the river current is 2 km/hr *relative to a point on the shore,* you conclude that your velocity relative to the shore is the *sum* of these two velocities,

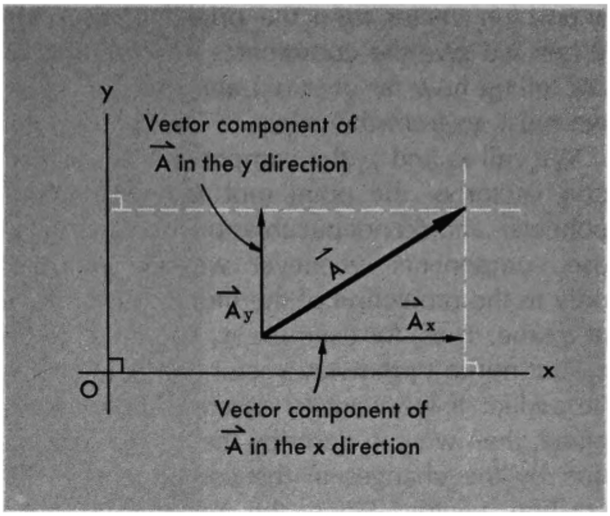


12–13 If you swim so that the rectangular component of $\vec{v}_w$ parallel to the banks is equal in magnitude but opposite in direction to $\vec{v}_c$, then $\vec{v}_b$ will be at right angles to the bank.

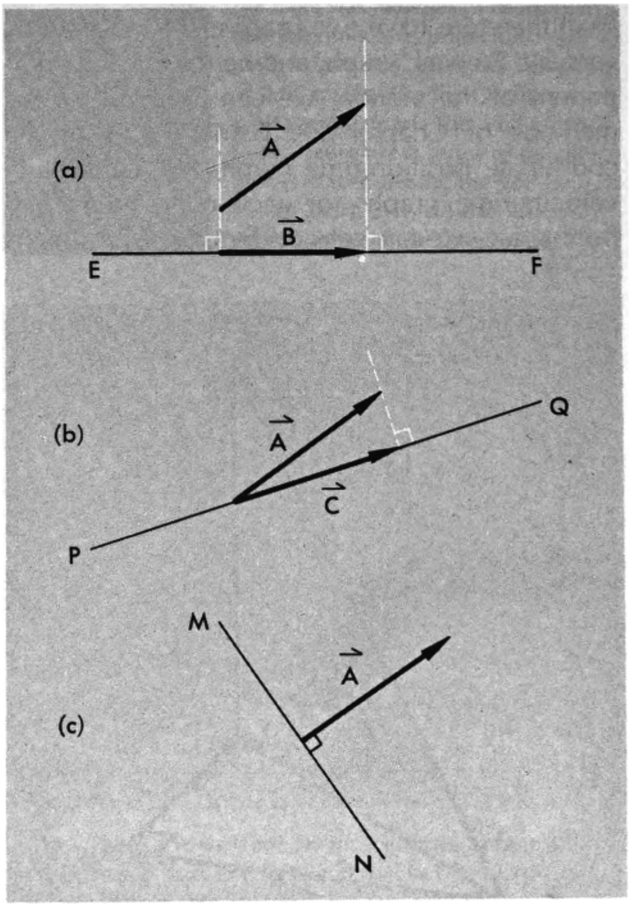
or 1 km/hr + 2 km/hr = 3 km/hr. Similarly, if you row at an angle with the stream, always heading in a direction of 45° with respect to the stream, you conclude that your velocity relative to a point on the river bank can be found by vector addition of the velocity vectors representing the river velocity relative to the shore and your velocity relative to the moving water. Any experiments you can do will yield velocity measurements that confirm the correctness of your conclusions when you deal with velocities, rivers, and boats. But, as you will learn in a later chapter, this simple vector addition of two velocities measured in *different* frames of reference —that is, velocities measured relative to different points moving with respect to each other—holds only approximately, and fails when the magnitude of one of them is a significant fraction of the speed of light. Moving objects such as cars, jet planes, rockets, and bullets do not move at anything like the speed of light relative to other objects on the surface of the earth, and we are safe in these cases in assuming that all velocities add like vectors.

In the case of the rising balloon, where the vertical velocity of the balloon and its horizontal velocity (the velocity of the wind) were both measured in the same frame of reference, our assumption is perfectly correct. It would even be correct if each of the two velocities were a significant fraction of the speed of light.

So far we have been discussing only vector components of the velocity vector. Actually, any vector $\vec{\mathbf{A}}$ on a plane surface can be represented by two rectangular vector components. First we choose two perpendicular reference directions—for example, the *x* and *y* coordinate lines as in Fig. 12 – 14. Then to construct the vector component of $\vec{\mathbf{A}}$ along the *x* axis, we construct a line through the head of $\vec{\mathbf{A}}$ perpendicular to the *x* axis. Now we draw the vector which extends parallel to the *x* axis, starting from the tail of $\vec{\mathbf{A}}$ and ending at the perpendicular. This gives the vector component along the *x* axis (sometimes called $\vec{\mathbf{A}}_x$, the vector component of $\vec{\mathbf{A}}$ in the *x* direction). To get the vector component in the *y* direction ($\vec{\mathbf{A}}_y$, the *y* vector component of $\vec{\mathbf{A}}$), we follow a similar process, constructing a perpendicular to the *y* axis, etc. (See Fig. 12 – 14.) In vector notation, $\vec{\mathbf{A}} = \vec{\mathbf{A}}_x + \vec{\mathbf{A}}_y$. Examples of some of the possible rectangular components of a vector are shown in Fig. 12 – 15.



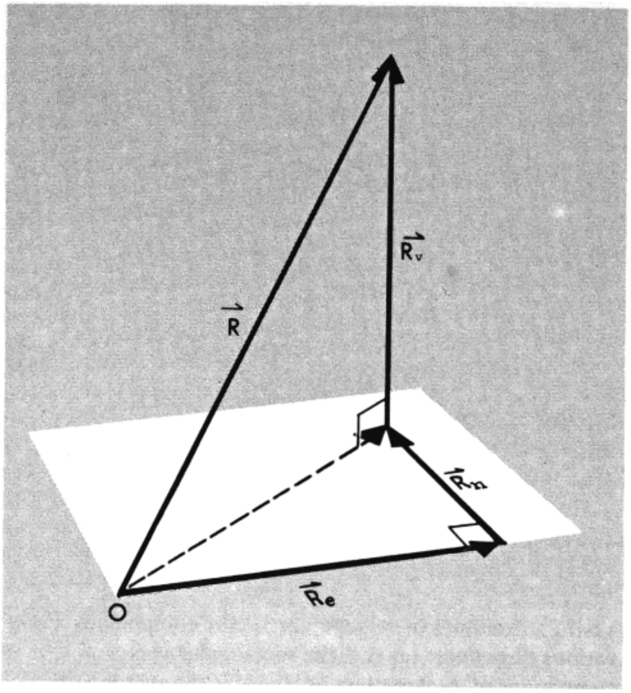
12–14 How to construct rectangular vector components in a plane. In vector notation, $\vec{A} = \vec{A}_x + \vec{A}_y$.



12–15 Examples of rectangular vector components along various directions. (a) $\vec{B}$ is the rectangular vector component of $\vec{A}$ along line EF. (b) $\vec{C}$ is the rectangular vector component of $\vec{A}$ along line PQ. (c) The rectangular vector component of $\vec{A}$ along line MN is zero.

In Section 12 – 1 we described the position of a point on a plane either by giving the two coordinate values of the point (x_1,y_1) or by drawing a position vector from the origin to the point. When we give the coordinates of the point, we are telling how far over (x_1) and how far up (y_1) we must go from the origin to locate the point.

We call x_1 and y_1 the components of the position vector of the point (*not* the vector components), and do not put an arrow over them. We use components whenever we are interested only in the projection of the motion on one axis at a time. Thus, for example, x_1 can be added to x_2, but not to y_1 the way vector components can be added. If we want to describe motion on a plane, then we can describe the change of position by the changes in the components of the position vector. Since the rectangular components are independent, as we saw in the case of the swimmer, we can treat each component separately as a case of straight-line motion. We shall therefore have two graphs: x versus t and y versus t. To each graph, and hence to each component of the motion, we can apply all of the methods of Chapter 11. Using slopes of the coordinate position-time graphs, we can obtain velocity-time graphs for each component, and from these we can get acceleration-time graphs for each component. We shall see an application of this to a specific problem in the next section.



12–16 $\vec{R}_e$, $\vec{R}_n$, and $\vec{R}_v$ are the east, north, and vertical vector components of vector $\vec{R}$, and $\vec{R} = \vec{R}_e + \vec{R}_n + \vec{R}_v$.

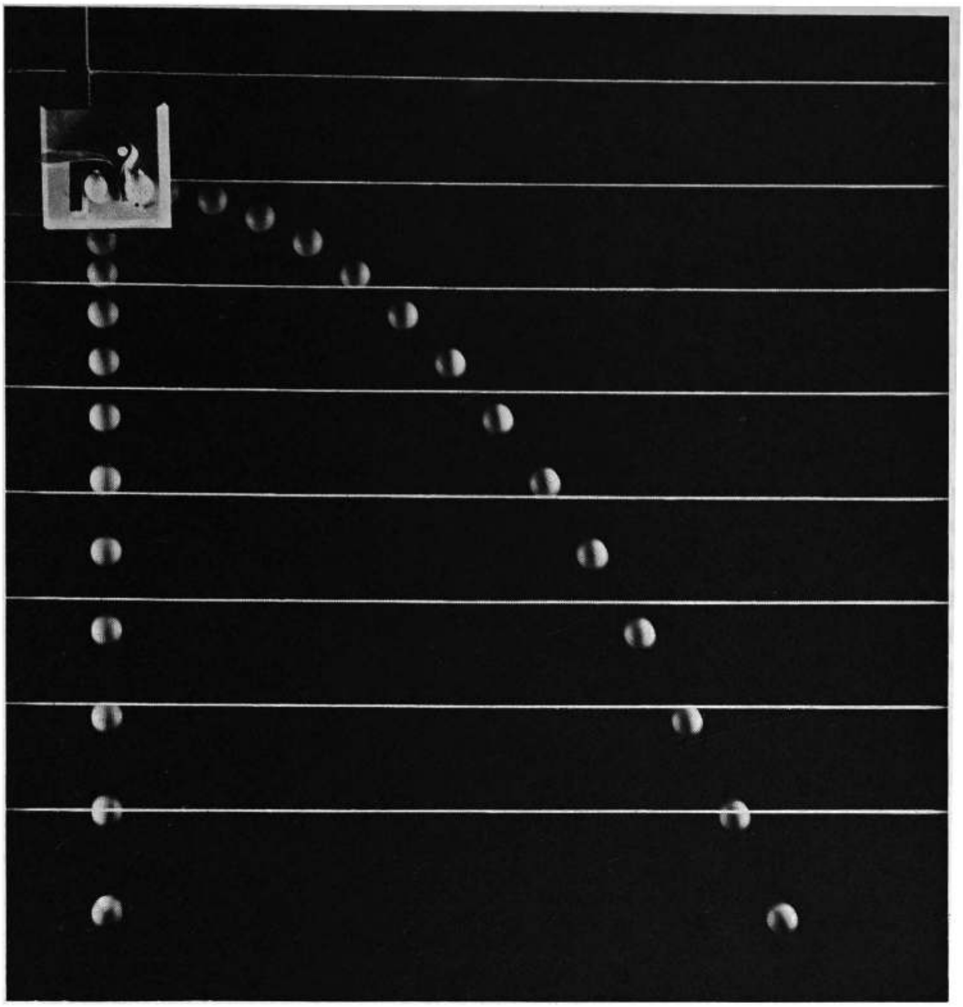
We have needed only two vector components to specify a vector because we have been considering vectors on a plane—that is, in two dimensions. In three dimensions we can specify a vector by giving its vector components along three mutually perpendicular directions. As illustrated in Fig. 12 – 16, these three rectangular components added together give the vector, $\vec{R} = \vec{R}_e + \vec{R}_n + \vec{R}_v$. To represent three-dimensional motion by the methods of Chapter 11 requires three coordinate vs. time graphs, but everything else goes along just the same. A velocity vector, for example, will have three components (v_x, v_y, v_z) where any of these three quantities can be either positive or negative. But again, v_x can be added only to another x component of a velocity and not to a y or z component.

12–6 VELOCITY CHANGES AND CONSTANT VECTOR ACCELERATION

Figure 12 – 17 is a multiple-flash photograph of two balls started in motion together. The ball on the left was allowed to fall straight down; the one on the right was projected horizontally with an initial velocity of 2.00 m/sec. Let us analyze the motions of both balls by finding their velocity vectors at successive intervals as they fall. We shall start with the ball on the left.

We can get the average velocity vector for a given time interval by measuring the distance between two images of the ball and dividing by the time between the flashes which made those images. This gives us the length of the average velocity vector; its direction is the direction of the displacement of the ball from one image to the other. In Fig. 12 – 18 we have measured off the successive displacements and put them side by side. The white strings in Fig. 12 – 17 are 15 cm apart and the interval between flashes is $\frac{1}{30}$ of a second. Using these facts, we have computed a scale so that we can read these vectors directly as the average velocities.

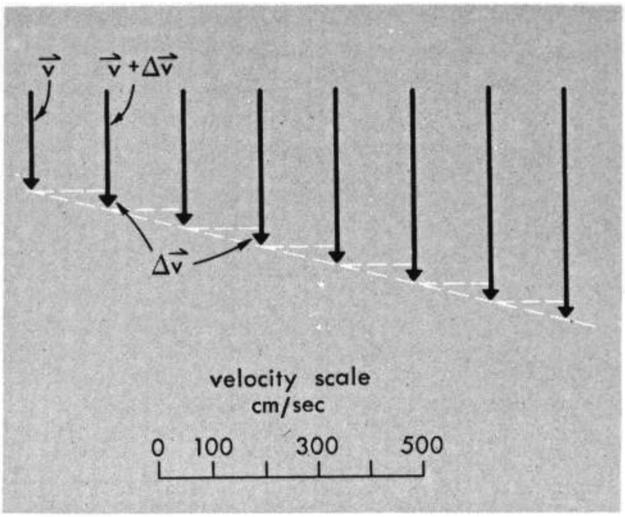
Figure 12 – 18 shows that the average velocity vector of the left-hand ball changes steadily. In each successive interval it increases by the same amount. The rate of change of the average



12–17 A flash photograph of two golf balls released simultaneously from the mechanism shown. One of the balls was allowed to drop freely, and the other was projected horizontally with an initial velocity of 2.00 m/sec. The light flashes were 1/30 second apart. The white lines in the figure are a series of parallel strings placed behind the golf balls, 15 cm apart. Why do the strings appear to be in the foreground?

velocity is therefore constant. Because the rate of change of average velocity is constant throughout the motion, we can be almost certain that the rate of change of instantaneous velocity is also constant throughout the motion and of the same size. To confirm this, of course, we might want to take additional strobe photographs of the falling balls, using shorter and shorter time intervals to approach the limit as $\Delta t \to 0$. But even without this, we can be pretty certain of our conclusion. Consequently, for the instant of the *n*th flash, we can write the instantaneous velocity as $\vec{v}_n = \vec{v}_i + n\Delta\vec{v}$. Here $\vec{v}_i$ is the instantaneous velocity vector with which we started and $\Delta\vec{v}$ is the constant change that occurs in each interval. By adding *n* of these changes to the original velocity, we get the velocity *n* intervals further along.

We can rewrite the last equation so that it more closely resembles the equations we developed for the description of motion along a straight-line path. (See Chapter 11, especially



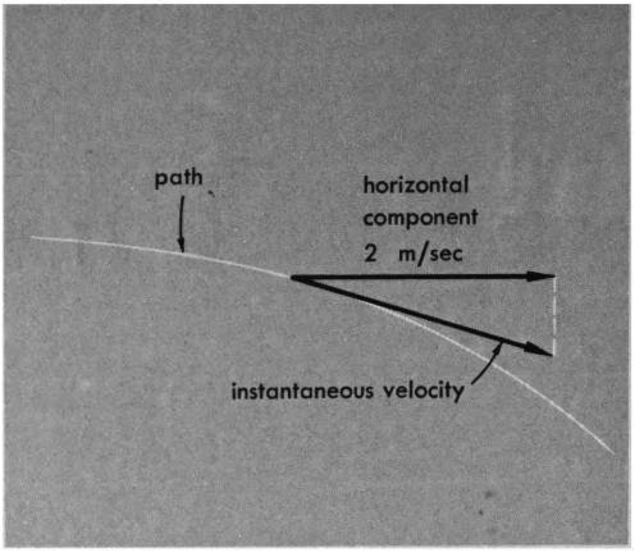
12–18 These arrows were drawn $2\frac{1}{2}$ times as long as the displacements of the left-hand ball in Fig. 12–17 during the last eight successive intervals of 1/30 of a second. Because we know the actual separations of the white lines in Fig. 12–17 and the time intervals, we can turn these arrows into average velocities for each time interval. The scale enables you to read the lengths of the arrows directly in units of cm/sec.

Sections 11 – 8 and 11 – 9.) There we defined the acceleration along the path as $a = \Delta v/\Delta t$. Here by dividing $\Delta\vec{\mathbf{v}}$ by Δt we shall introduce the vector acceleration $\vec{\mathbf{a}} = \Delta\vec{\mathbf{v}}/\Delta t$. Using it, our last equation becomes

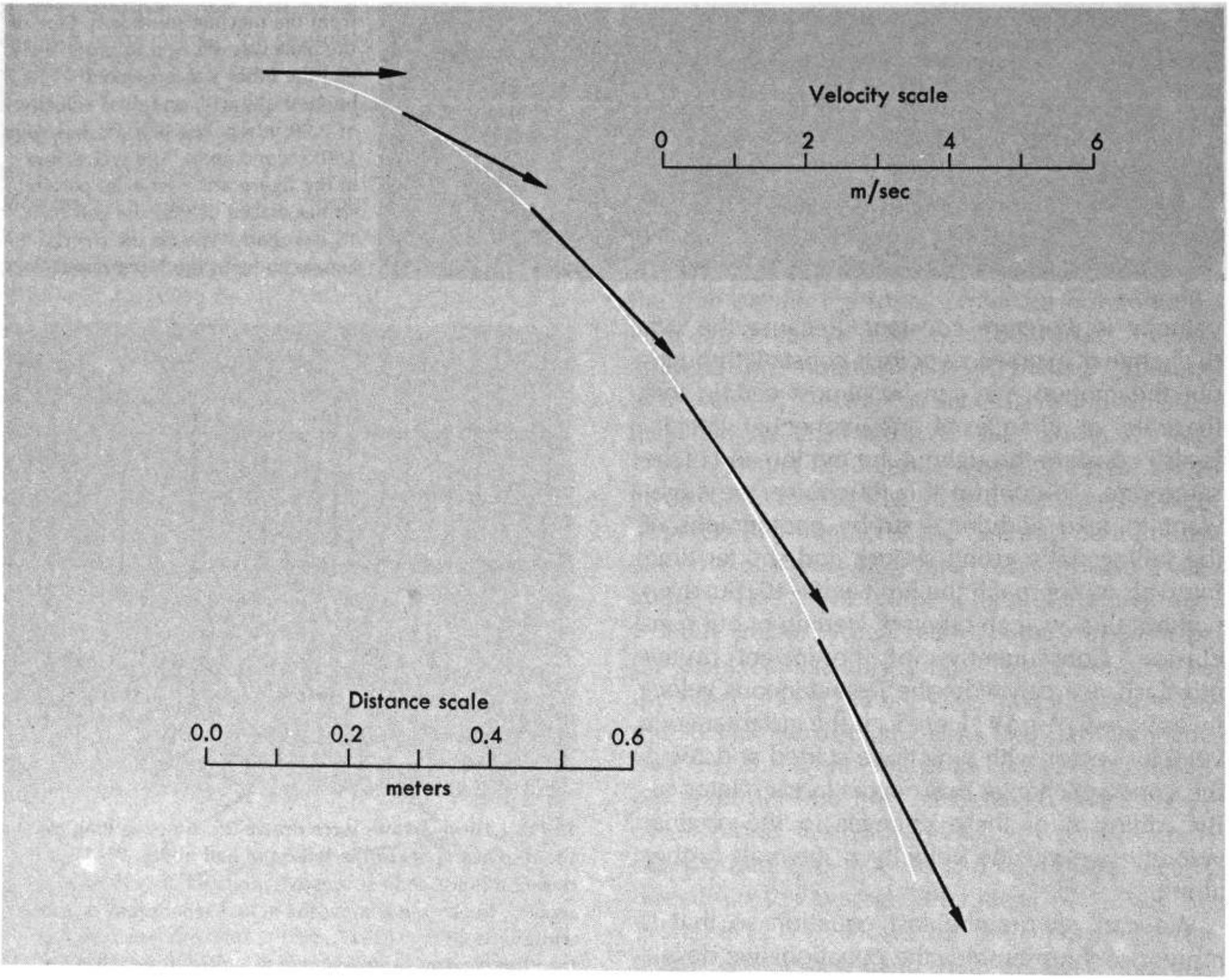
$$\begin{aligned}\vec{\mathbf{v}}_n &= \vec{\mathbf{v}}_i + n\Delta t\frac{\Delta\vec{\mathbf{v}}}{\Delta t}\\ &= \vec{\mathbf{v}}_i + \vec{\mathbf{a}}\,t\end{aligned}$$

In the last line of this equation, we have replaced $n\Delta t$ by t, the time during which the velocity has changed from its initial value $\vec{\mathbf{v}}_i$ to its value when the nth flash occurs.

We have just described the motion of a falling ball in vector language. But since the ball moves on a straight (downward) path, we hardly need the vectors. We could have done the analysis equally easily with the methods of Chapter 11. The vector language, however, becomes far



12–19 How to find the instantaneous velocity vector. It is tangent to the path and of such length that its horizontal component is equal to the initial horizontal velocity of the projectile.



12–20 The position and velocity of the "thrown" golf ball in Fig. 12–17 are shown here on a single graph.

more useful when we analyze a more complicated motion. To see this, let us get back to Fig. 12 – 17 and study the motion of the other ball, the one that moves out to the right in the figure.

The second ball in Fig. 12 – 17 moves both to the right and down. From the fact that the distance between the positions of the ball at successive flashes of the strobe light is greater for the later pictures, we see that the speed is increasing. Since the path is not a straight line, the direction of the velocity is changing too. We can analyze Fig. 12 – 17 to get the instantaneous velocity of the ball at various points along the path.* The results of such an analysis are shown in Fig. 12 – 20. There both the position of the ball and its instantaneous velocity at 0.10-sec intervals are shown in the same graph. Note the two scales, one for distance and the other for velocity.

Figure 12 – 21 shows only the sequence of velocity vectors of Fig. 12 – 20. Here, however, we have drawn the velocity vectors from the same starting point. Examination of this figure shows us that the successive vectors are obtained by adding a velocity vector of about 1 m/sec (actually it is 0.98 m/sec) directed vertically downward. We can express this rule in equation form. To do this, we first express the components of the velocity. The horizontal component of velocity

$$v_h = 2.00 \text{ m/sec, to the right}$$

stays constant throughout the flight. On the other hand, the vertical component is zero at $t = 0.00$ sec and increases in magnitude by 0.98 m/sec during each 0.10 sec. This is a uniform change at the rate of -9.8 m/sec^2, so the vertical component of velocity at any time t is downward and is

$$v_v = (-9.8 \text{ m/sec}^2)t$$

where t is the time in seconds.

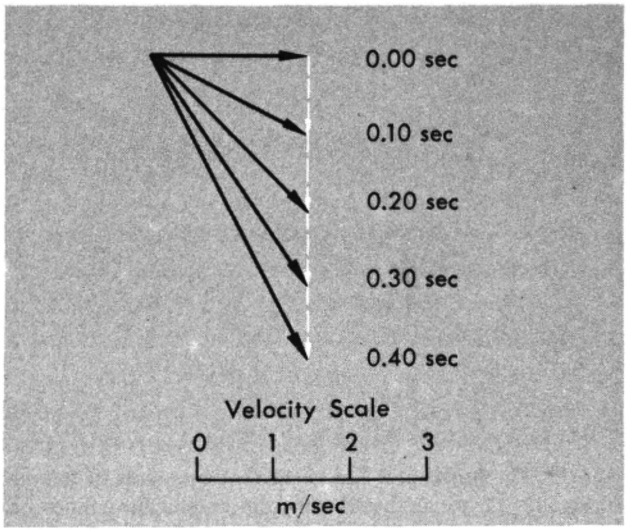
With the two rectangular velocity components identified, we can now use the graphical methods of Chapter 11 to find the components of the acceleration and displacement vectors. These are shown in Fig. 12 – 22.

We can also use these velocity components to express the motion entirely in vector language. Combining the two vector components, the vector velocity at any time t is

$$\vec{\mathbf{v}}\,t = 2.00 \text{ m/sec, to the right} + (9.8 \text{ m/sec}^2)t\text{, downward}$$

We can put our equation for the motion of the ball thrown to the right into just the same form as we did for the ball falling straight down. Here (2.00 m/sec, to the right) is the initial velocity $\vec{\mathbf{v}}_i$, and (9.8 m/sec², downward) is the constant acceleration $\vec{\mathbf{a}}$. So we get $\vec{\mathbf{v}}_t = \vec{\mathbf{v}}_i + \vec{\mathbf{a}}\,t$ again. Notice that the acceleration of both balls is down. Also the downward motion is the same for both, as you can see by looking across the picture to check that one moves down the same amount as the other in each time interval. The only thing different in describing the motions of the two balls is the value of $\vec{\mathbf{v}}_i$. The acceleration vector $\vec{\mathbf{a}}$ is the same for both.

In Chapter 11 we found that $v_t = v_i + at$ described the velocity at any time for straight-line motion in which there was constant acceleration along the path. The equation $\vec{\mathbf{v}}_t = \vec{\mathbf{v}}_i + \vec{\mathbf{a}}\,t$ describes the velocity vector at any time for motion along any path, as long as the acceleration vector is constant in both magnitude

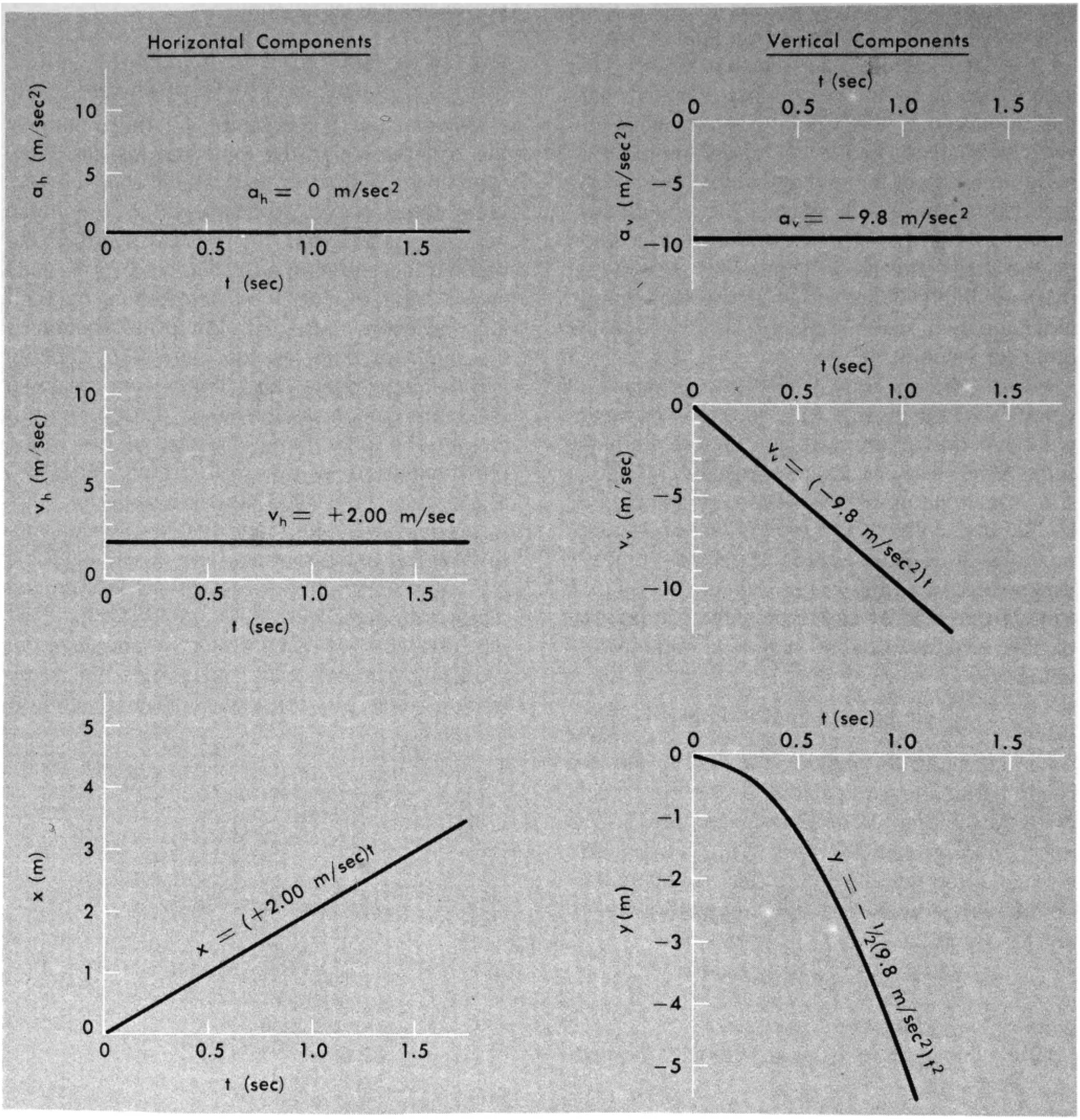


12–21 A sequence that shows only the velocity vectors of Fig. 12–20. Successive vectors are found by adding a constant vector directed vertically downward.

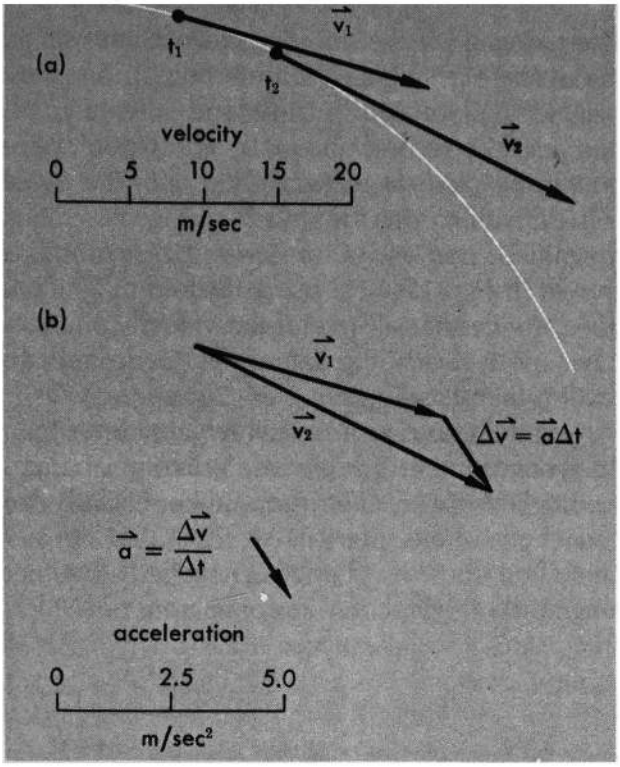
---

*One way of finding the instantaneous velocity in this case is to note that the horizontal component of the velocity is constant. This follows from the fact that the horizontal displacement is the same in each time interval. We then get $\vec{\mathbf{v}}$ from this fact and the fact that the instantaneous velocity vector always points in the direction of the path (See Fig. 12 – 19.)

and direction. Notice that the acceleration vector need not point along the path, as in the example of the right-hand ball in Fig. 12 – 17. Furthermore, $\vec{a}$ can be any constant vector of the right units. The -9.8 m/sec² which we found for the balls, is just a special value that occurs when balls move freely near the surface of the earth. In studying other motion, we shall find other constant values of $\vec{a}$ – and also acceleration vectors that change as time goes on. For instance, by pushing a ball we can get almost any $\vec{a}$ we wish.

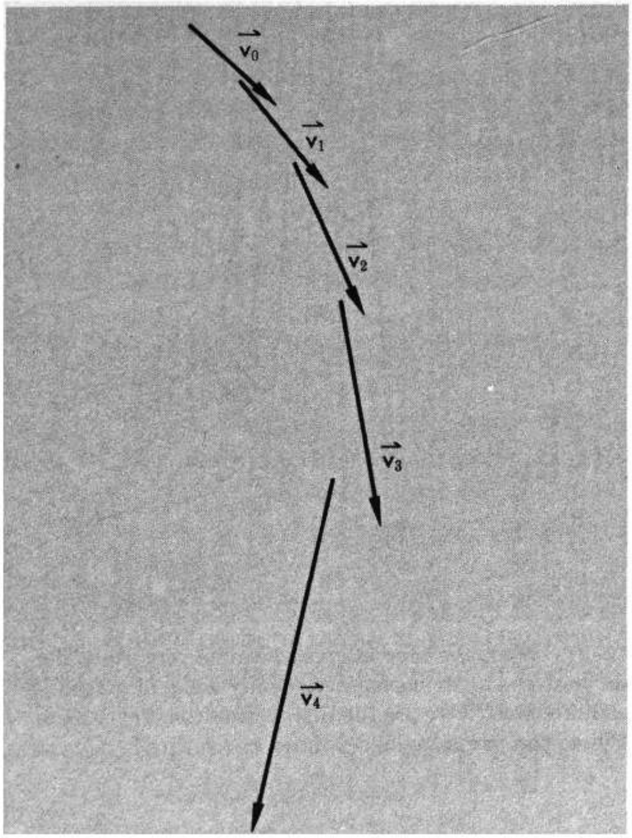


12–22 The horizontal and vertical components of the motion of the righthand ball of Fig. 12–17 are analyzed using the graphical methods of Chapter 11. The positive x coordinate line is to the right; the positive y coordinate line upward. As functions of time, we find that the horizontal component of position is $x = (2.00 \text{ m/sec})\, t$; the vertical component of position is $y = \frac{1}{2}(-9.8 \text{ m/sec}^2)\, t^2$. If we eliminated t from these two equations, we would obtain an equation for the path in space, y as a function of x.



12–23 To find the average acceleration in the interval $\Delta t = t_2 - t_1$, first find the vector difference $\Delta\vec{v} = \vec{v}_2 - \vec{v}_1$ and divide this by Δt. The result is the average acceleration vector $\vec{a}$, which can now be plotted with an appropriate scale as shown above.

12–24 Successive velocity vectors of the speedboat at 10-second time intervals.



12–7 CHANGING ACCELERATION

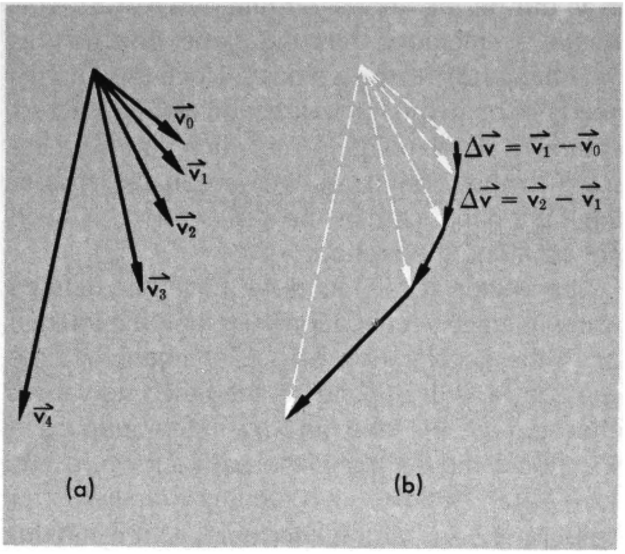
In the last section we described motion with constant vector acceleration. We introduced the vector acceleration to describe how the velocity vector changes. Even when the acceleration vector is not itself constant, we can introduce it in just the same way. We define it by $\vec{a} = \Delta\vec{v}/\Delta t$, where $\Delta\vec{v}$ is the vector change in $\vec{v}$ during the very short time interval Δt. This vector acceleration has the same direction as the change $\Delta\vec{v}$ of the velocity; since this change need not be in the same direction as $\vec{v}$, the acceleration $\vec{a}$ may point in any direction with respect to the motion. As we saw in the case of the right-hand ball in the last section, it need not be along the motion.

The vector acceleration we have just defined is the average acceleration over the time interval Δt. If the acceleration is itself changing as time goes on, $\vec{a}$ will depend on the time interval we choose. Let us take an example. Suppose a speedboat moves along the path shown in Fig. 12 – 23 (a). At time t_1, it is moving with the vector velocity $\vec{v}_1$, and at the later time, t_2, it is moving with vector velocity $\vec{v}_2$. What is the average acceleration in the time interval $\Delta t = t_2 - t_1$? To find out we must determine the vector change $\Delta\vec{v} = \vec{v}_2 - \vec{v}_1$ in the velocity. Then the average acceleration is $\vec{a} = \Delta\vec{v}/\Delta t$. The procedure is indicated in Fig. 12 – 23 (b). We take the vector difference between $\vec{v}_2$ and $\vec{v}_1$, divide $\Delta\vec{v}$ by Δt, and plot the average acceleration vector $\vec{a}$ for the time Δt on a new diagram with an appropriate scale.

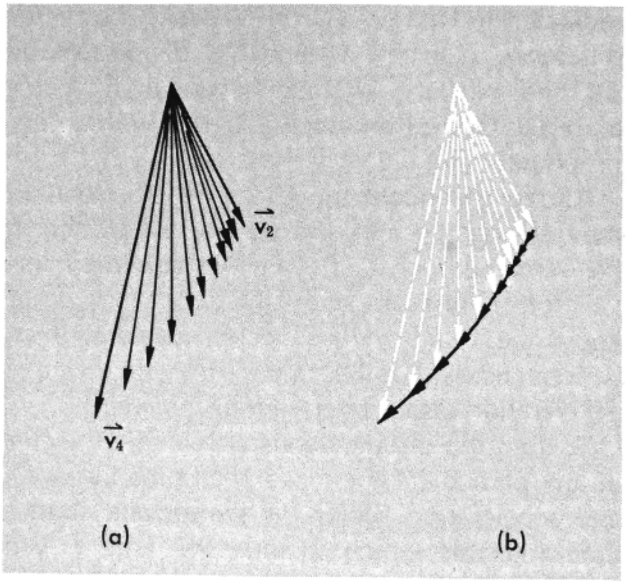
As the speedboat moves on, its acceleration may change. In fact, we do not expect that it has remained constant even over the time interval from t_1 to t_2. In such situations we need to know the instantaneous vector acceleration at various times t rather than only the average acceleration vector over various intervals.

To find the instantaneous acceleration vector at any particular time t, we determine the average acceleration vector for shorter and shorter time intervals which include the time t. We define the instantaneous acceleration at a given time as the limit of the average acceleration $\Delta\vec{v}/\Delta t$ as the interval Δt becomes smaller and smaller. Usually this limiting vector has a definite size and direction.

Suppose, for example, that Fig. 12 – 24 shows the successive velocity vectors of our speedboat at time intervals of 10 seconds. If we make a composite picture in which the tails of these vectors coincide, we get Fig. 12 – 25. The changes in velocity during each successive time interval are also shown in this figure. We see that the velocity changes $\Delta\vec{v}$ in each of the four intervals differ both in magnitude and in direction. Therefore, since the time intervals are the same, the average accelerations are different.


$\vec{v}_0$
$\vec{v}_1$
$\vec{v}_2$
$\vec{v}_3$
$\vec{v}_4$
$\Delta\vec{v} = \vec{v}_1 - \vec{v}_0$
$\Delta\vec{v} = \vec{v}_2 - \vec{v}_1$
(a)
(b)

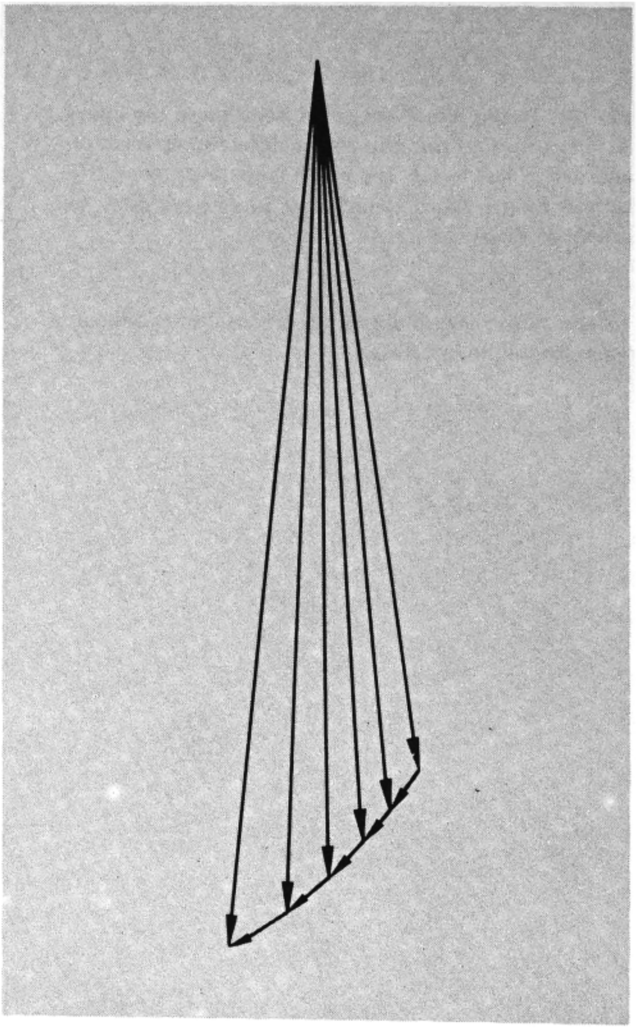
12–25 The vectors of Fig. 12–24 are shown here drawn from the same origin. The acceleration is not constant in direction or magnitude, as shown in (b).



12–26 If we take 2-sec time intervals, the motion described in Fig. 12–24 looks like this when the vectors are drawn from the same origin. By taking a shorter time interval, the changes of velocity become more nearly equal.

If, instead of using 10-second intervals, we had used 2-second intervals, we would have gotten the pictures shown in Fig. 12 – 26. Since it is difficult to see the details of Fig. 12 – 26, a magnified portion of the first 10 seconds is shown in Fig. 12 – 27, using fivefold magnification. Notice that all five of the velocity changes have more nearly the same magnitude and direction than those of Fig. 12 – 25 (b).

If we had chosen a time interval shorter than 2 seconds, the successive velocity changes would have been even more alike. Usually we can choose time intervals so short that the acceleration does not change appreciably in either magnitude or direction in going from one inter-



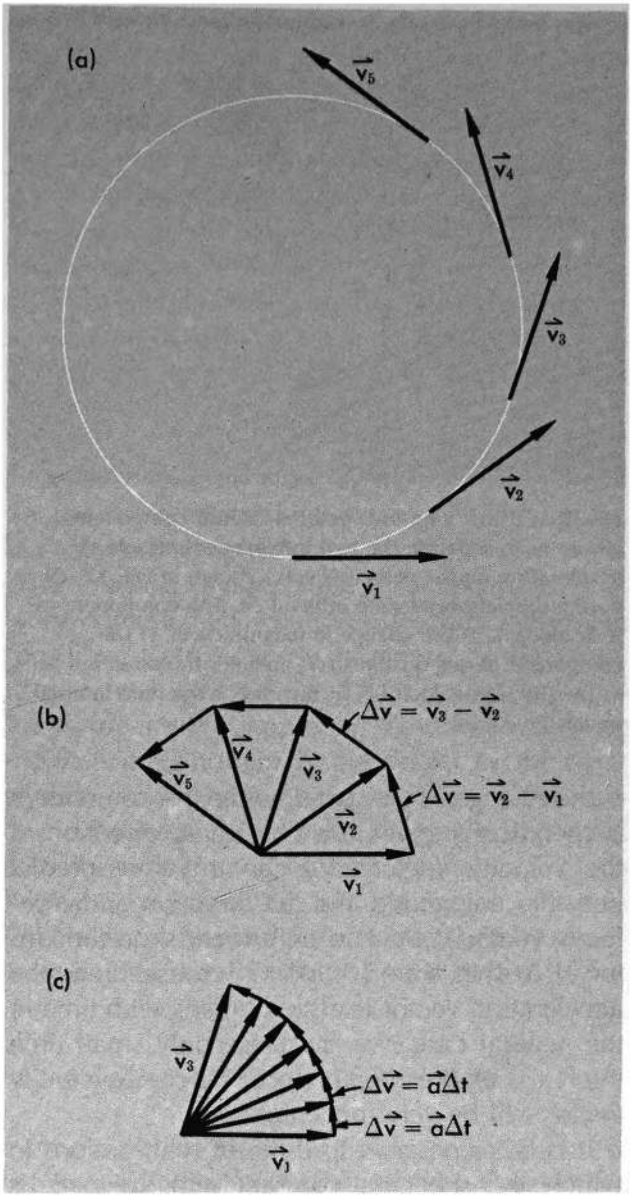
12–27 When the time intervals become very short, the successive velocity changes are nearly equal in magnitude and direction. Here are the first 10 sec from Fig. 12–26. The vectors are magnified 5 times.

val to the next. With this understanding of "very short," we see that the instantaneous vector acceleration at time t is the average acceleration for a very short time interval that includes the time t. We can therefore say that the instantaneous acceleration is the limit of the ratio $\Delta\vec{\mathbf{v}}/\Delta t$ as Δt becomes very small:

$$\vec{\mathbf{a}} = \lim_{\Delta t \to 0} \frac{\Delta\vec{\mathbf{v}}}{\Delta t}$$

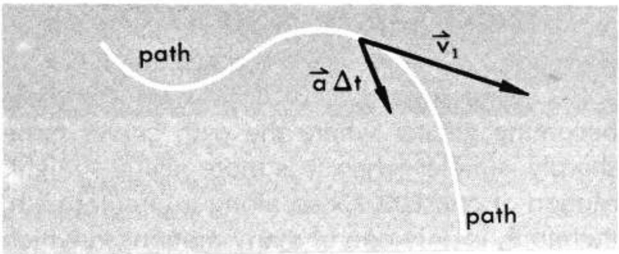
The really new thing that we learn by considering vector acceleration as compared with acceleration along a straight line is that *changes in the direction* of the velocity indicate that there is an acceleration. Even if a body moves along a curved path at constant speed, it is accelerated. The simplest and most important example of this is a body moving in a circular path at constant speed. This is accelerated motion and the acceleration is a changing one. In Fig. 12 – 28 the circular path is shown and on it a sequence of velocity vectors at equal time intervals. These vectors are all of the same length, but each points in a direction different from any other. If we construct a velocity diagram, as in Fig. 12 – 28 (b), in which the velocity vectors are plotted from a common origin, we see that the successive changes $\Delta\vec{\mathbf{v}}$ in velocity are also all of the same length but different in direction one from the other. Because the successive changes in velocity are not parallel to each other, the vector acceleration which measures the change of velocity with time cannot be constant. By taking smaller and smaller time intervals, as in Fig. 12 – 28 (c), we can see that the instantaneous acceleration vector is directed perpendicular to the velocity vector at each instant of time. The fact that the magnitude of the acceleration vector is constant *and* the angle between the acceleration and velocity vectors is always 90° means that they rotate at the same uniform rate.

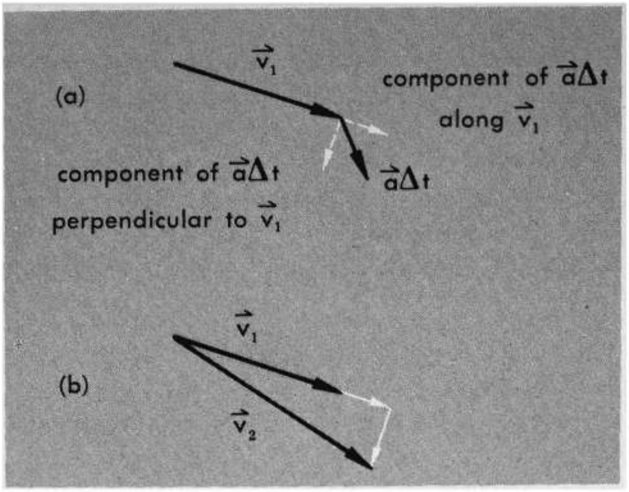
In the general case, when a body moves on a curved path and the speed is not constant, the acceleration vector will stand at some angle to the velocity vector at any instant. This is shown in Fig. 12 – 29. If the acceleration vector is broken into two components, one along the path and the other perpendicular to the path as in Fig. 12 – 30 (a), we can then separate their effects on the velocity during a very short time interval Δt. The component of $\vec{\mathbf{a}}\,\Delta t$ along the path


(a)
$\vec{v}_5$
$\vec{v}_4$
$\vec{v}_3$
$\vec{v}_2$
$\vec{v}_1$
(b)
$\Delta\vec{v} = \vec{v}_3 - \vec{v}_2$
$\vec{v}_4$
$\vec{v}_5$
$\vec{v}_3$
$\Delta\vec{v} = \vec{v}_2 - \vec{v}_1$
$\vec{v}_2$
$\vec{v}_1$
(c)
$\vec{v}_3$
$\Delta\vec{v} = \vec{a}\Delta t$
$\Delta\vec{v} = \vec{a}\Delta t$
$\vec{v}_1$

12–28 Part (a) shows the velocity of a body moving at constant speed in a circular path. In (b) the velocity vectors are drawn from a common origin, showing that the changes in velocity are in different directions. When shorter time intervals are taken, as in (c), the instantaneous acceleration vector is seen to be perpendicular to the velocity vector.

12–29 The acceleration vector stands at some angle to the velocity vector in the case of motion along a curved path at nonconstant speed.




(a)
$\vec{v}_1$
component of $\vec{a}\Delta t$
along $\vec{v}_1$
component of $\vec{a}\Delta t$
perpendicular to $\vec{v}_1$
$\vec{a}\Delta t$
(b)
$\vec{v}_1$
$\vec{v}_2$

12–30 (a) Breaking the vector $\vec{a}\,\Delta t$ into components allows us to separate the two independent effects of acceleration on the velocity vector shown in Fig. 12–29 during the very brief time interval Δt. The component of $\vec{a}\,\Delta t$ along $\vec{v}_1$ is the change in magnitude of $\vec{v}$; the component perpendicular to $\vec{v}_1$ changes the direction of $\vec{v}$. (b) The new velocity $\vec{v}_2$ at the end of the time interval equals $\vec{v}_1 + \vec{a}\,\Delta t$.

is the change in the speed, while the component perpendicular gives the change in direction of the velocity vector. We can therefore predict both the magnitude and the direction of the velocity vector at the end of the very short time interval Δt [Fig. 12 – 30 (b)]. Of course, since the acceleration vector is also changing with time in the general case, we must use only small time intervals, or the prediction of the new velocity vector will be very inaccurate.

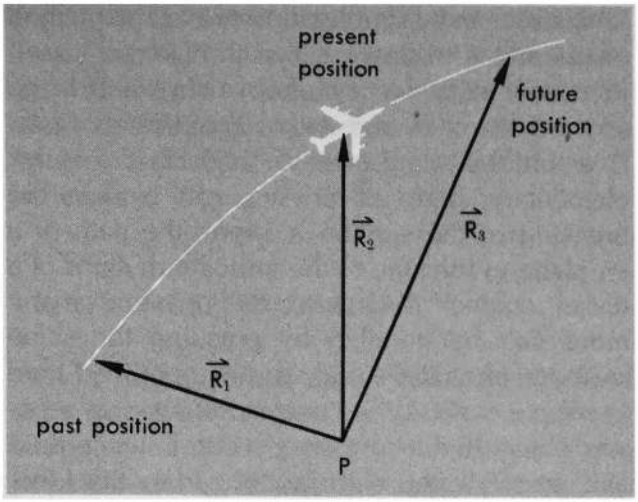
It is quite possible to think of real motions in which the speed is constant and the path is curved but not necessarily circular; the speedboat discussed at the beginning of this section, for example, could satisfy these conditions. What can we say about the acceleration-velocity relations, and how do they compare with the case for circular motion at constant speed? Because the speed does not change, the instantaneous acceleration then can have no component along the path. The acceleration vector is perpendicular to the velocity vector; that is, perpendicular to the path at all points. However, the magnitude of the acceleration vector will change from point to point on the path, becoming greater where the path curves more sharply, smaller when it is more nearly straight. Motion at constant speed along a circular path, therefore, is only one of many motions in which the velocity vector and the acceleration vector are perpendicular to each other. It is the special case in which the magnitude of the acceleration is constant.

12–8 THE DESCRIPTION OF MOTION; FRAMES OF REFERENCE

We have studied vectors largely in order to describe motion, to describe the successive positions of an object in space and how fast it moves through them. The graceful motions of a waltz can be described fully, if somewhat dully, as a sequence of positions of the hands, the feet, and the rest of the bodies of the dancers, an appropriate position for each instant of time. In fact, dancing to music implies a manner of time measurement. For each note, there is a correct figure of the dance; and to each note, a proper instant of time.

The simplest and most fundamental motion is the motion of a single object, whose parts we do not find it necessary to distinguish, either because we cannot observe them or because they remain fixed relative to one another. A planet, such as Jupiter, moves in the sky. We can represent its position at each instant of time by a vector which we imagine drawn from the earth to the planet Jupiter. An airplane drones on its course toward some destination; its positions can also be represented by vectors (Fig. 12 – 31). Neither the planet nor the airplane is a tiny point; and, if we wish to know what happens within them, we would need much more information than is given by these position vectors. But for many of the needs of astronomy or of navigation, it is sufficient to think of each as a point at the tip of a vector. As the planet or airplane moves, the vector moves, changing in magnitude and direction.

In each of our examples the position vector extends from some identifiable point called the origin to the moving object, from the earth to Jupiter, from the control tower to the airplane. Furthermore, motion is always measured with respect to some frame of reference, and so are positions. The airplane's motion is described with respect to the earth's surface, for example, and so is the motion of the falling balls in Fig. 12 – 17.



12–31 The positions of an airplane with respect to a point may be shown by position vectors giving distance and direction.

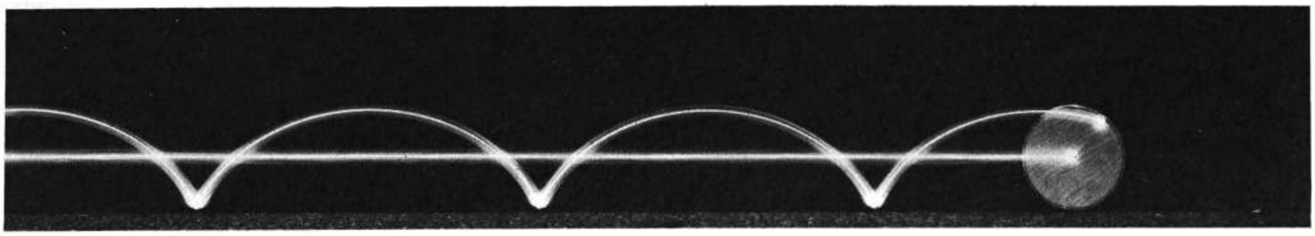
The motion of the airplane with respect to another airplane may be very different from its motion with respect to the ground. The motion of the thrown ball in Fig. 12 – 17 is just motion with a constant velocity with respect to the other ball falling on the left side of the figure. The motion of Jupiter with respect to the sun is simpler than its motion with respect to the earth. To the driver of a moving car a raindrop falling vertically with respect to the earth rushes almost horizontally toward him. Motion is described differently depending on the frame of reference with respect to which we give the description.

In general, we wish to view motion in such a way as to make it appear simple. We therefore place ourselves mentally in a frame of reference in which the motions are easy to describe, and draw the position vectors locating the object from the most convenient point. Another example will show what we mean. Suppose we stand on the earth again and look at the motion of a point on the edge of a slowly moving wheel of a car. The point moves through the curve illustrated in Fig. 12 – 32, a complicated curve known as a cycloid. A position vector from us as origin to a point on the wheel performs an extremely complicated motion. But a point on the earth is not always the most convenient origin for position vectors. We are far better off if we get in the car and hang out of the window to look at the wheel. The point on the wheel then moves steadily around a circle, and the motion of the position vector looks far simpler.

In describing motion we try to put the origin in the most convenient place even though we are sometimes physically unable to go there. The motions of the planets, for example, look very complicated when we describe them with position vectors whose origin is in the earth. Copernicus pointed out that the planetary motions are much simpler to describe when we assume that the sun is the center of the solar system, and move the origin of the position vectors to the sun. Ever since Copernicus pointed out the importance of placing the origin in a convenient place, the proper choice of origin has been an important technique used by physicists to describe motion in simple terms.

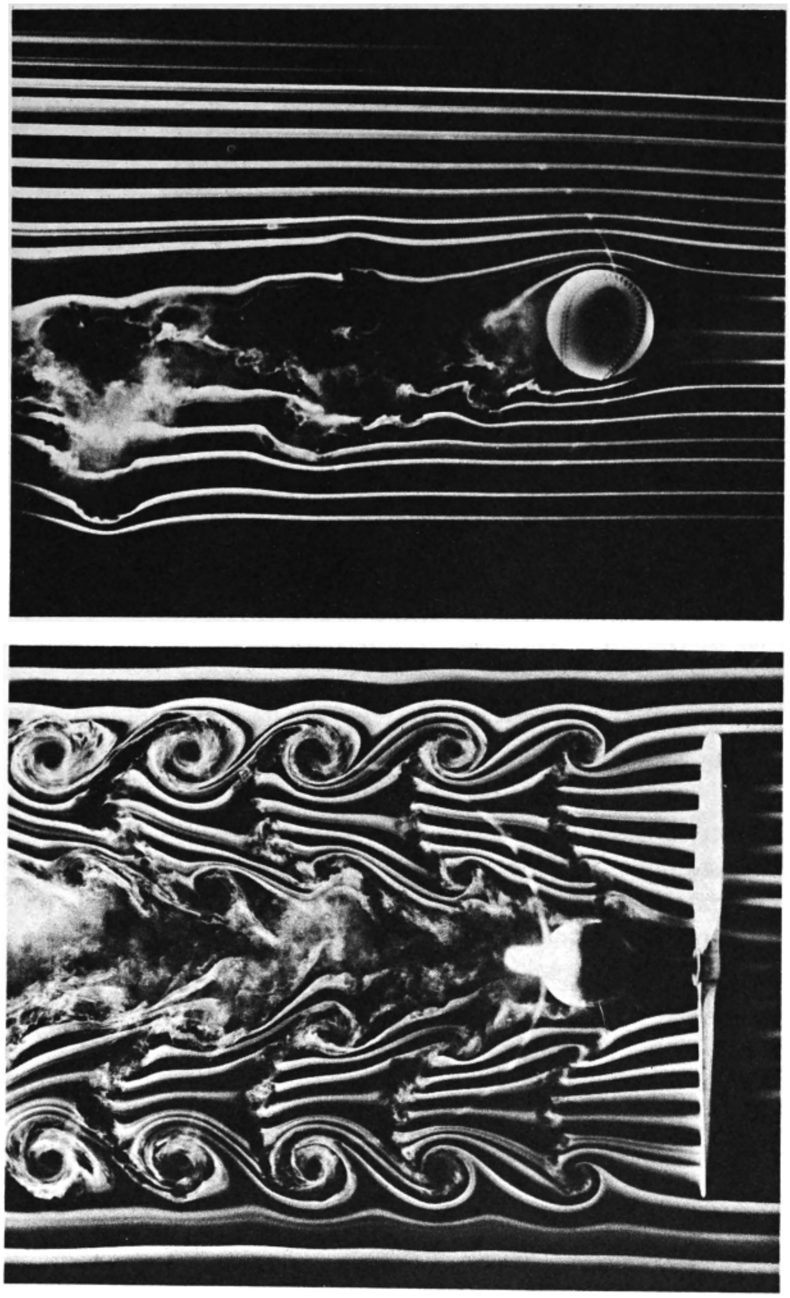
12–9 KINEMATICS AND DYNAMICS

What we have been discussing is the branch of physics called *kinematics* (after the Greek word *kinema*, meaning "motion"). It is that part of our subject which treats of the description of motion, without taking into account what it is that moves or what causes the motion. The science of mechanics, indeed much of physics, is dominated by the study of motion; but that



12–32 Path of a point on the rim of a rolling wheel. This curve, a cycloid, shows the path as it appears to an observer standing alongside. One small light was mounted on the rim of a wheel, and another light at the center. The camera shutter was held open while the wheel rolled.

study is complete only when we extend it to what is called *dynamics* (after the Greek word *dynamis,* meaning "power"). In dynamics one discusses the causes of motion, what is moving, and how its nature affects the motion. In our study of kinematics, we needed to measure only positions and times; in dynamics, the pushes and pulls which cause and resist and determine motions must be taken into account as well.



12–33 (Above) The motion imparted to air particles by a rapidly spinning ball, such as a pitcher throws for a curve, is an example of a very complex form of motion. (Below) Eddies formed behind a propeller spinning at 4080 rpm. This is another example of a highly complex motion. (Courtesy: F. N. M. Brown.)

Motion can be simple or complex. The method of physics is to analyze first the simpler cases, to extract what we can from them, and to go forward to more and more complicated cases. It would be a mistake to think that a quick elementary study of physics can explain the breaking of the surf on a shore, the path of a jet plane in the sky, or the intricate rhythms of a diesel engine. But these motions and many more can be handled by pursuing the same methods of analysis that we have outlined here in simple cases. When, instead of a single separate object like a pebble or a car, a whole fluid like air or water is in motion, even the kinematics may become difficult. The eddies behind a spinning propeller and a thrown baseball (Fig. 12 – 33) are complex indeed. Their study is a specialized kind of unraveling, requiring enormous "bookkeeping" operations to keep track of every small bit of the moving air. Powerful mathematical methods are needed, but the basic ideas used in solving such problems flow out of the discussion we have given; the underlying laws are not different. One may well be wary, though, of too simple an explanation of what obviously is kinematically complex. You cannot design airplanes or understand the weather without checking theories with experiments or without interpreting experimental results in terms of theories. We must go back and forth between theory and the experimental study of the real, complex motions. Nevertheless, great progress has been and will be made with such a process. Nowhere in the whole domain of the motions of the natural world, here on earth or in the skies, in our large-scale machines or in our means of transport, have we yet found any case to which the analysis of mechanics does not apply. Where the kinematics is hard, progress is slow; where the kinematics is less demanding, progress is fast. In this course, where we are trying to learn fundamentals, we do not want to produce the false impression that there are no complex cases of interest; there are, especially in technology. In engineering and technology, the aim is to get something done for people to use; in a science such as physics, the aim is to understand and hence to predict and to control. In physics we seek out the simple and examine it, even if it is costly and difficult to attain. In engineering the orderly methods developed in this research are painstakingly and often brilliantly applied to more and more complex cases, which have clear usefulness for men. The complicated, trial-and-error extension always rests on the foundation of the simple and the well understood.

12–10 THE SPEED OF LIGHT

We have pointed out earlier that from everyday experiences we can develop a wide extension or extrapolation. This is the kind of thing we have just been doing, basing our study of motion on the ideas of space and time we have built up from experience. But if such schemes are pushed too far, they may go wrong. They need to be tested. For fifty years now, we have been finding that our usual notions of space and time, extremely reliable for most of the motions that we notice daily, do not work for extremely fast motions. For speeds which begin to be of the order of magnitude of that of light, the kinematics we have just discussed begins to go wrong. It turns out that the speed scale cannot be extended indefinitely; if you add velocities that are too big, you reach a region where the rules of addition go wrong. There arises a natural limit to speed, which cannot be surpassed. This universal speed limit is the speed of light in free space, which is known to be very close to 3×10^8 m/sec. The speed of the fastest rockets does not exceed about 2×10^4 m/sec. Speeds up to 10^5 m/sec span the whole range of speeds of every large object in the solar system, from planets and meteors to the engineering devices of man. Only small particles, electrons and their kin, move appreciably faster. These closely approach the speed of light. For their study, the ideas of relativity kinematics are needed. But for everything else—solar-system astronomy, engineering, or any large-scale laboratory physics, the kinematics we have studied, that of Newton, is accurate enough. This whole topic is an example of how you can begin with familiar ideas, which hold well over a wide range, and reform them entirely when you reach another order of magnitude. The fact that familiar ideas may be modified to fit extreme conditions does not end their meaning and usefulness in the domain for which they were originally built up and in which they have been amply tested. Newtonian kinematics is a good approximation to relativity kinematics whenever the speeds are small compared with the speed of light.

QUESTIONS

1.* You may walk the following distances, one after another, in any order you choose: (a) go 3 meters due east; (b) go 2 meters due north; (c) go 3 meters due west. What is the farthest you can be from your starting point at the end of your walk?

2. Show that the sum and difference of two perpendicular vectors of equal lengths are also perpendicular and of the same length.

3. A man follows this route: From his house he travels four blocks east, three blocks north, three blocks east, six blocks south, three blocks west, three blocks south, two blocks east, two blocks south, eight blocks west, six blocks north, and two blocks east. How far and in what direction will he be from home?

4. Suppose you have three sticks 0.40, 0.50, and 0.60 m long, which can be joined by connecting the ends at a 90° angle.

   (a) What is the largest displacement possible between the unconnected ends of the sticks?

   (b) Displacement vectors can be added in any order and the result is the same. Why do you get different lengths between the ends of the sticks in this problem when you change the order of the sticks?

5.* What is the difference between (5)(6 km/hr, west) and (5 hr)(6 km/hr, west)?

6. (a) Find the result of adding a vector 2 cm east to one 3 cm northwest.

   (b) Find the result of adding a vector 8 cm east to one 12 cm northwest.

   (c) Compare the results of parts (a) and (b), and state a theorem about adding a pair of vectors which are multiples of another pair. Can you prove the theorem in general?

7.* If you are in a free balloon being carried along by the wind at a constant velocity, and you hold a lightweight cloth flag in your hand, which way will the flag wave?

8. Suppose an airplane flies in a circle of circumference 20 km at a constant speed of 200 km/hr.

(a) What is the change in velocity in one fourth of a revolution?

(b) What is the change in velocity in one half of a revolution?

9. An airplane maintains a heading of due south at an airspeed of 900 km/hr. It is flying through a jet stream which is moving east at 400 km/hr.

(a) In what direction is the plane moving with respect to the ground?

(b) What is the plane's speed with respect to the ground?

(c) What distance over the ground does the plane travel in 15 min?

10. An airplane is flying toward a destination 400 km due east of its starting point, and the wind is from the northwest at 60 km/hr. The pilot wishes to make the trip in 40 min.

(a) What should be the heading?

(b) At what airspeed should he fly?

11. A steamer is traveling directly south at 25 km/hr in an area where the wind is from the southwest at 18 km/hr What is the angle from true north of the smoke trail from the stack?

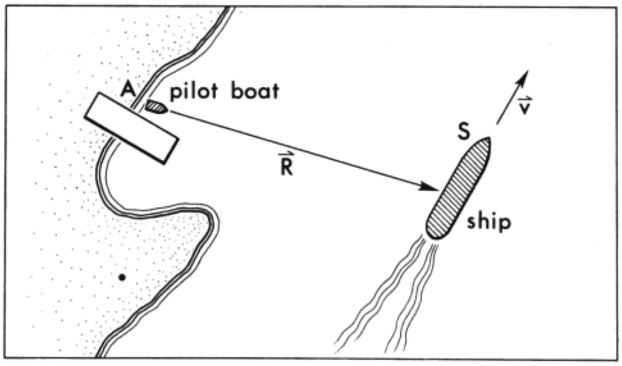
12. A man rows a boat "across" a river at 4.0 km/hr (i.e., the boat is always kept headed at right angles to the stream). The river is flowing at 6.0 km/hr and is 0.20 km across.

(a) In what direction does his boat actually go relative to the shore?

(b) How long does it take him to cross the river?

(c) How far is his landing point downstream from his starting point?

(d) How long would it take him to cross the river if there were no current?



12–34 For Question 13.

13. Suppose you are the captain of a pilot boat about to leave from a harbor shown at *A* in Fig. 12 – 34 and you wish to meet a ship *S* which is $\vec{R}$ miles away from you and steaming at $\vec{v}$ km/hr. It is a calm day with a negligible wind and no current, and your boat has a speed of *u* km/hr with respect to the water.

(a) What direction would you steer to meet the ship?

(b) Where will you meet the ship?

(c) How could you determine how long it would take to meet the ship?

14 * If north and east are selected as perpendicular reference directions, what are (a) the vector components and (b) the components of a velocity vector of magnitude 100 km/hr and direction 30° west of south?

15.* If the *x* and *y* axes are perpendicular reference directions, what is the vector component in the *y* direction of a vector of length 2 in the *x* direction?

16. What is the magnitude of a displacement whose components along the perpendicular *x*, *y*, and *z* axes are respectively 4.00 m, 2.50 m, and 8.50 m?

17. The components of two separate vectors are (x_1, y_1) and $(-x_1, -y_1)$ Draw a diagram showing the vector components of these two vectors.

18.* Suppose you were told that Fig. 12 – 17 represented two balls thrown *upward*. In what way would Fig. 12 – 21 have to be changed in order to correspond to the motion of the right-hand ball?

19. (a) Draw a log-log graph of the vertical displacement as a function of time of the right-hand golf ball in Fig. 12 – 17.

(b) What would a log-log graph of the total displacement from the origin as a function of time for the right-hand ball look like?

20.* Is motion at constant speed restricted to either a straight line or a circular path?

21.* How can a motion with constant speed be an accelerated motion?

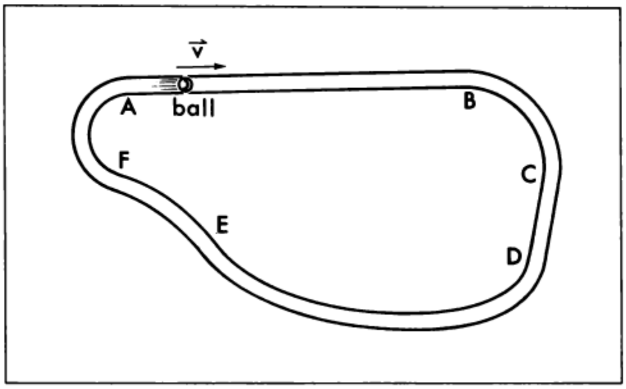
22. A ball is rolling in a channel that is in a horizontal plane (Fig. 12 – 35). If the friction is negligible,

(a) will the speed of the ball change?

(b) will its velocity change?

(c) what is the direction of the acceleration?

(d) is the magnitude of the acceleration always the same?



12–35 For Question 22.

23. An object moving in a circular path with a constant speed of 2.0 m/sec changes direction by 30° in 3.0 seconds.

(a) What is its change in velocity?

(b) What is its average acceleration during the 3.0 seconds?

24. A watch has a second hand 2.0 cm long.

(a) Compute the speed of the tip of the second hand.

(b) What is the velocity of the tip of the second hand at 0.0 seconds? at 15 seconds?

(c) Compute its change in velocity between 0.0 and 15 seconds.

(d) Compute its average vector acceleration between 0.0 and 15 seconds.

25.* Describe the motion of a swing as it is seen from another swing exactly like it, when both swings started together in the same direction with the same push.

26.* In Fig. 12 – 17, how does the left-hand ball appear to move, to a fly sitting on the other ball?

27. A bug placed at the center of a phonograph turntable crawls at a constant speed directly toward a point on the edge of the turntable. Describe the motion of the bug as seen by you from above when the turntable has a speed of 16 revolutions per minute and it takes the bug 15 seconds to crawl from the center to the edge of the turntable.

28. The radius of Mars' orbit about the sun is about 1.5 times the radius of the earth's orbit, and the period of Mars' revolution around the sun is about twice that of the earth.

(a) Draw the "solar system" of these two planets and the sun to scale and construct displacement vectors from Earth to Mars that are about 1/2 earth-month apart for two revolutions of the earth.

(b) Use the above displacements to draw the path of Mars relative to the earth.

(c) How often, during a two-year period, does Mars appear to reverse its direction of motion as seen from the earth?

29. A plane flying north at 320 km/hr passes directly under another plane flying east at 260 km/hr.

(a) What is the horizontal vector component of the displacement of the second plane relative to the first 20 minutes after they pass each other? 50 minutes after they pass?

(b) What is the horizontal vector component of the velocity of the plane flying east relative to the plane flying north?

(c) Does the direction of this velocity vector (relative to the earth) change?

30. An airplane takes off due south from an airport and maintains a ground speed of 4.0 km/min while climbing at a rate of 0.30 km/min. After 0.50 min, it turns due east, maintaining the same ground speed and the same rate of climb. At the end of 1.0 min after taking off:

(a) How high is the airplane?

(b) What is its direction from the airport?

(c) How far is the point on the ground directly under the airplane from the point of take-off?

(d) How far is the airplane from the take-off point?

31. A navy pilot, in plotting his trip to rendezvous with his carrier, learns that he is 600 nautical miles due south of the carrier at the time his position is determined. The carrier is on a course 15° east of north and is making 25 knots (nautical miles per hour). The wind at his flight altitude is from 70° west of north at 40 knots. What must be his heading and air speed to make his rendezvous in just two hours' time?

32. An ocean liner is traveling at 18 km/hr. A passenger on deck walks toward the rear of the ship at a rate of 4.0 m/sec. After walking 30 meters he turns right and walks at the same rate to the rail, which is 12 meters from his turning point.

(a) What is his velocity relative to the water surface while walking to the rear? While walking toward the rail?

(b) Draw the displacement vectors relative to the water surface for his stroll. What was the total displacement from his starting point?

13

Newton's Law of Motion

AUTOMOBILES move down highways or wind in and out of traffic; passenger planes fly high above us; jet planes and artificial satellites streak across the sky; the stars perform their regular progression. What makes each of them go? What makes anything move? Is there a single cause common to all motion? Is any cause necessary?

So far in our study of physics we have been concerned almost entirely with building up a systematic way of describing and analyzing physical phenomena. We have measured the sizes of atoms, the distances to stars, and the lengths of time intervals; we have learned to recognize the angles in reflection and the patterns of interference of waves. In the study of motion this systematic description of observations is called *kinematics* (Chapter 11). It is the description of motion without regard to the causes of the motions. But description alone can never enable us to satisfy our desire to do something new, to control motions, to go beyond the mere description of what occurs. In this part of the book we shall take the next step: we shall examine the causes of motions or of changes in motion. This study is called *dynamics*.

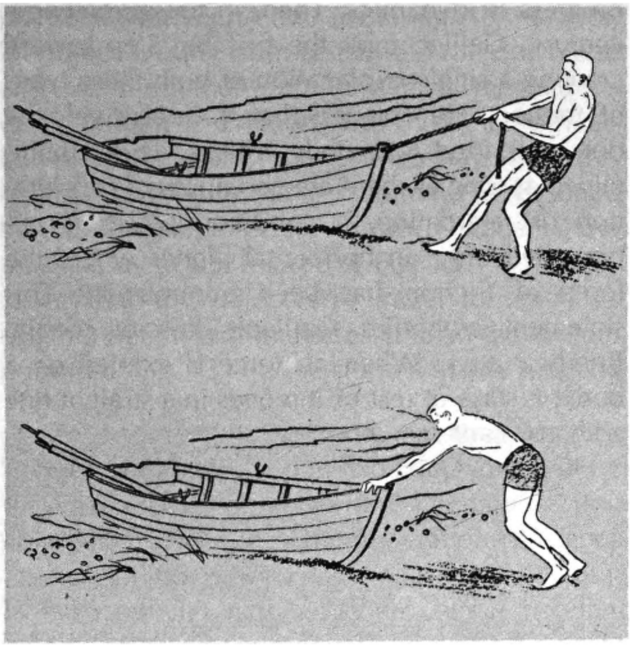
Newton's law of motion, on which we now base our understanding of dynamics, does go beyond kinematic description. For example, we make use of it when we design rockets and launch man-made satellites. In this part of the book, once we have understood Newton's law, we shall apply it to the motions of the moon and planets. Like Newton, we shall find the connection between the time it takes a planet to move around the sun and the gravitational attraction between chunks of matter. Later we shall use the same law of motion to study electrical forces and to enter the submicroscopic world. With this single law we shall investigate motion throughout all this range.

Amid the complex motions of complicated bodies—flying, tumbling, and vibrating—we shall find certain aspects of simplicity. With the aid of Newton's law we can find in all these various motions some properties which remain unchanged, new quantities which are conserved while all else changes. The conservation laws eventually carry us outside the range of mechanical behavior. We can change mechanical energy into electrical energy. But to understand energy we start with Newton's law.

13–1 IDEAS ABOUT FORCE AND MOTION

Questions about the causes of motion arose in the mind of man more than twenty-five centuries ago, but our present answers were not found until the time of Galileo (1564 – 1642) and Newton (1642 – 1727).

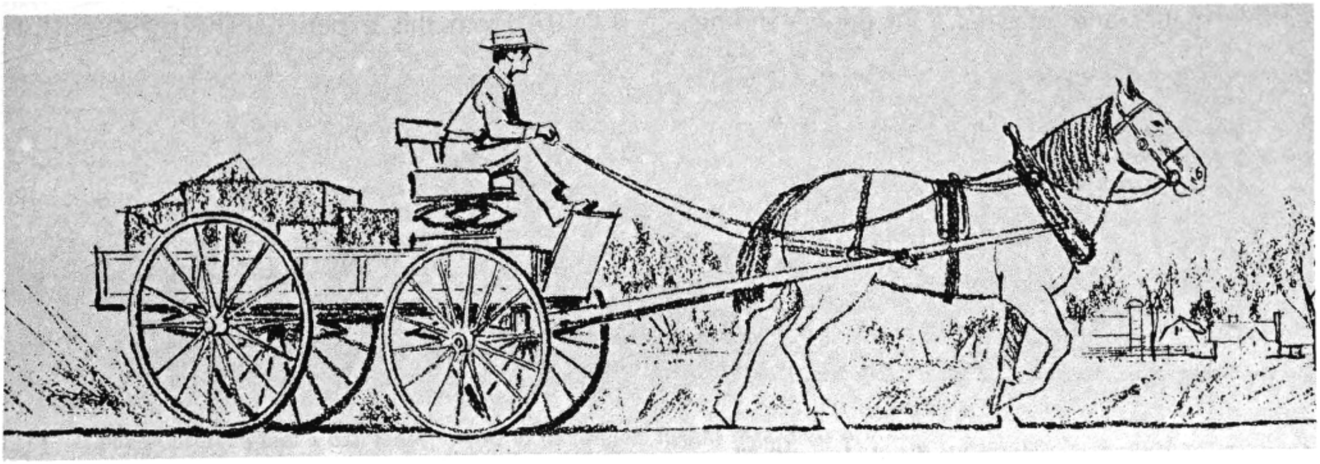
Let us start in terms of our own personal experience. What sort of thing do we associate with the "cause of motion"? The answer is muscular pulls or pushes (Fig. 13 – 1). To move a piano across a room you have to do a lot of pushing. To move a sheet of paper off your desk takes very little push. These pulls or pushes we call *forces*. The notion of force as used in physics certainly started this way. Later, as understanding grew, the idea of force was extended to include all causes of motion. The pull of a magnet on a nail is a force; it can change the nail's motion in the same way that a muscular force can.



13–1 All pushes and pulls are called forces.

More specifically, what is the relation between force and motion? Suppose we move a desk across the floor We must apply a force all the time to keep it moving steadily from one side of the room to the other. Similarly, a horse must keep pulling on a wagon to keep it rolling at constant speed. Everyday experience seems to indicate that it is necessary to exert a force constantly in order to maintain a steady motion, such as motion in a straight line with constant speed (Fig. 13 – 2). Aristotle (384 – 322 B.C.) had noted this fact. He concluded that a constant force was required to produce a constant velocity. It then follows that bodies would come to rest in the absence of force.

The hypothesis that in the absence of outside forces bodies would come to rest and stay at rest helps us to understand a great many observed motions, but it does not explain *all* the motions which occur in nature. For example, the Greeks were aware that bodies fall with increasing speed without the application of any evident outside force. They were also acquainted with the motions of the sun, moon, and stars, which seem to occur without pushes or pulls to maintain them. There seemed to be three kinds of motion. We must explain not only the motion of things we push around on the surface of the earth, but also the motion of bodies falling to earth and the unceasing motions of heavenly



13–2 Steady motion seems to require a steady force.

bodies. Aristotle explained that ordinary matter falls toward the earth because the earth is the center of the universe to which matter naturally moves. He proposed that celestial matter was fundamentally different in nature from matter on earth and that it therefore obeyed different laws. To Aristotle, celestial matter had the built-in property of supplying from within itself the force necessary to maintain the observed motions.

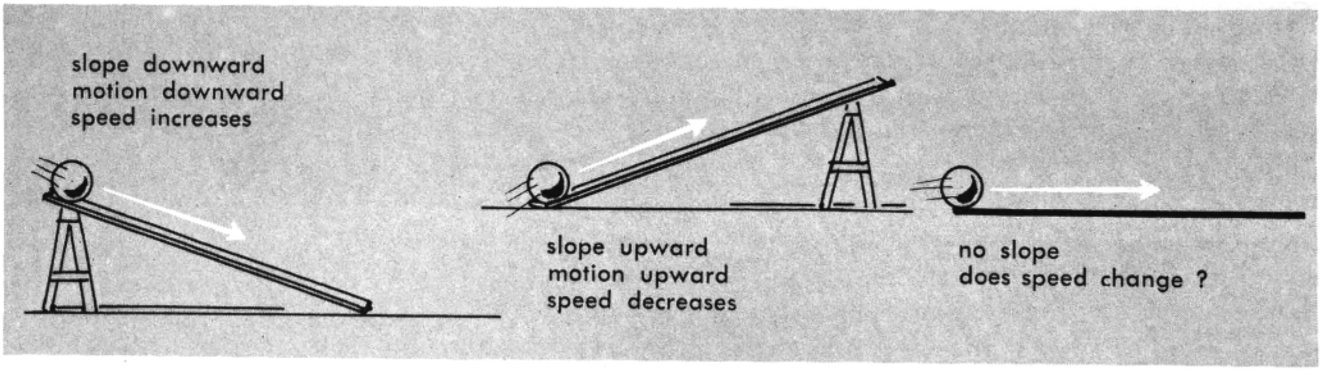
We need not think that these separate explanations for three different classes of observed motions are foolish. We often do the same thing. When we see a piece of metal that attracts iron nails, we say it is a magnet—a different kind of matter from wood; and we may investigate its magnetic behavior separately from its non-magnetic behavior. When we see a comb that attracts our hair, we say it is electrified, and we may explain its electrical behavior separately from its usual mechanical behavior. Of course we try, as the Greeks tried, to explain all we observe, but there are other desirable objectives. Explaining as much as we can with as few assumptions as possible is preferable to making a separate model for each new observation. As far as we can, we describe wood, magnets, and electrified combs in a single model, as simple as we can make it. Likewise, we try to explain all motion on one theory rather than three.

A modern Aristotle would hardly explain the unceasing nature of celestial motion by invoking a distinct kind of matter. We can send our own earthly matter into the celestial realm. The world of motion on earth and the unceasing motions of the planets are now united. The artificial satellites offer us an excellent demonstration that we need assume no difference between earthly matter and celestial. Our understanding of the motion of falling bodies, of heavenly bodies, and of bodies which we ourselves push and pull along the surface of the earth is now described in a single fundamental law of motion. The satellites were designed, built, and fired according to this law. Their behavior is one of many pieces of evidence that Newton's law of motion encompasses the three types of motions described by Aristotle.

13–2 MOTION WITHOUT FORCE

For two thousand years after the time of Aristotle, the apparent difference between celestial motion and motion on earth halted significant progress in dynamics. Then, in the seventeenth century, Galileo took the first big step toward creating a single explanation of both these types of motion. He asserted that ". . . any velocity once imparted to a body will be rigidly maintained as long as there are no causes of acceleration or retardation, a condition which is approached only on horizontal planes where the force of friction has been minimized." This statement embodies Galileo's law of inertia. Briefly it says: When no force is exerted on a body, it stays at rest or it moves in a straight line with constant speed.

How did Galileo reach the startling conclusion, so different from everyday experience, that constant motion requires no force? He was studying the motions of various objects on an inclined plane. He noted that "in the case of planes that slope downward there is already present a cause of acceleration, while on planes sloping upward there is retardation." (See Fig. 13 – 3.) From this experience he reasoned that



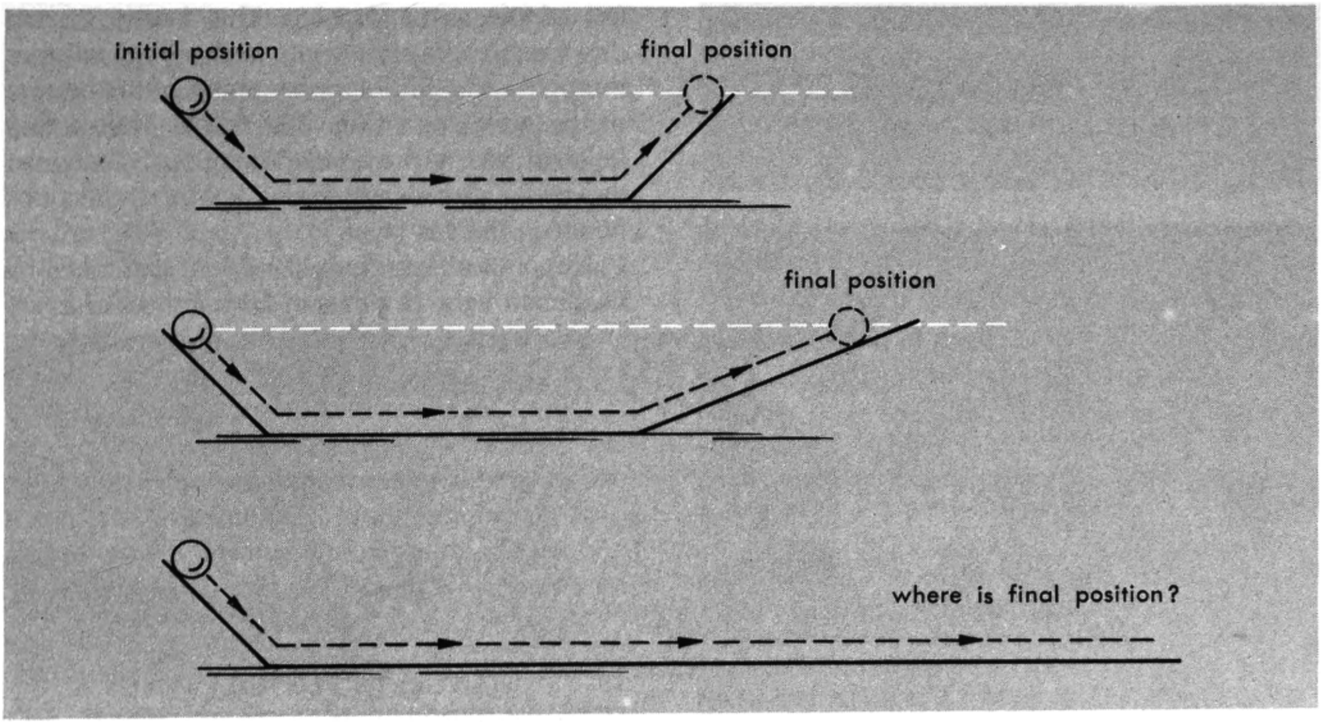
13–3 **From observing motion on inclined planes, Galileo reasoned that motion along a horizontal plane is steady.**

when the plane slopes neither upward nor downward there should be neither acceleration nor retardation: ". . . motion along a horizontal plane should be constant." Of course Galileo knew that such horizontal motions were not in fact constant, but he saw that when there was less friction, bodies moved for a longer time with nearly constant velocity. Because of his arguments he was convinced that friction provided the forces which stop bodies in horizontal motion, and that in the absence of all forces the bodies would continue to move forever. He therefore stated his result for the idealized situation in which *no* forces act.

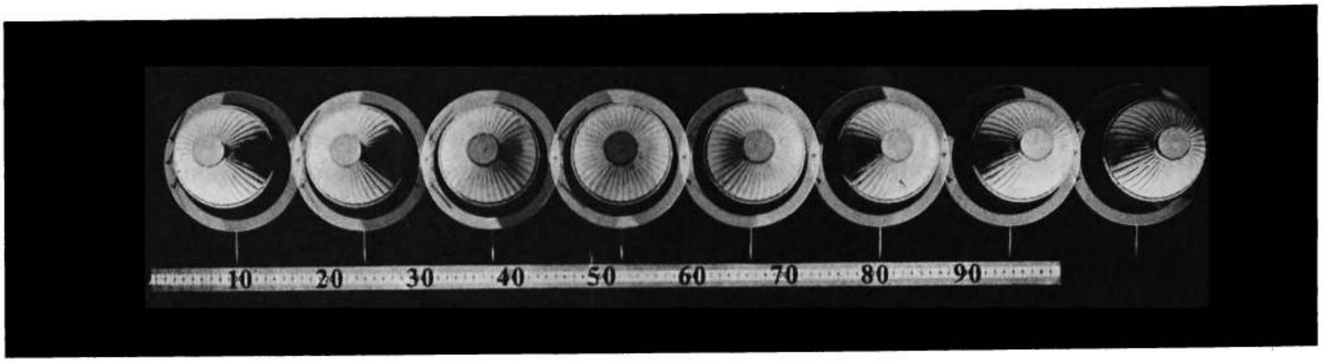
In a second series of experiments, Galileo showed that if he placed two of his inclined planes facing each other (as in Fig. 13 – 4, top), an object starting from rest would roll down one and up the other until it almost reached its original height. Friction prevented it from attaining this height, but Galileo saw that this height was the limit to the motion. He reasoned that if the slope of the upward plane is decreased, as in the middle of Fig. 13 – 4, the distance that the object will travel to reach its original height will increase. If, as in the bottom portion of Fig. 13 – 4, the slope if finally reduced to zero, so that the second plane is a horizontal surface, the object will never attain its initial height. It should travel on forever, "From this," Galileo again concluded, "it follows that motion along a horizontal plane is perpetual."

Galileo's experiments are not difficult, nor is there any evidence that he performed them with exceptional skill. Some, like the extension of the experiment at the bottom of Fig. 13 – 4 to the idealized case of perpetual motion, were not "real" experiments. They were experiments in thought alone. But they were based on solid fact. It is just this combination of thinking and fact that distinguishes Galileo's work. It was this combination which allowed him to pick out the useful idealization despite the great variety of observed motions. His principle of inertia was the great breakthrough which enabled Newton to build up our present understanding of dynamics.

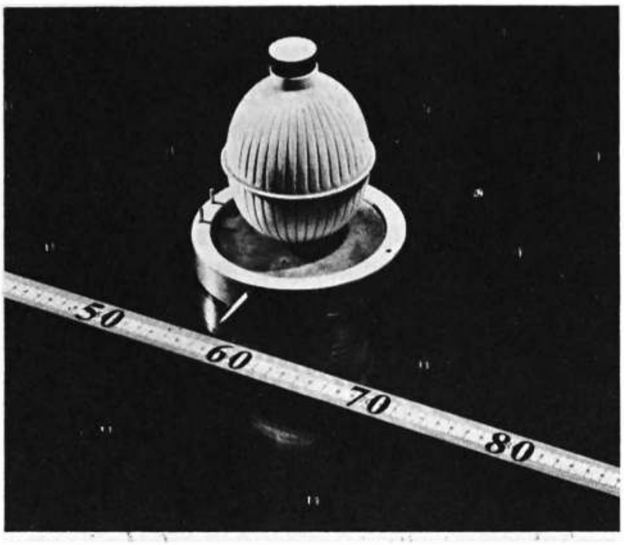
Many of the motions Galileo analyzed, and those which Newton later studied, were so highly idealized that they seem to have very little to do with motions of real systems as we observe them. But it was only by careful consideration of these idealized situations that Galileo and Newton made their great contributions to me-



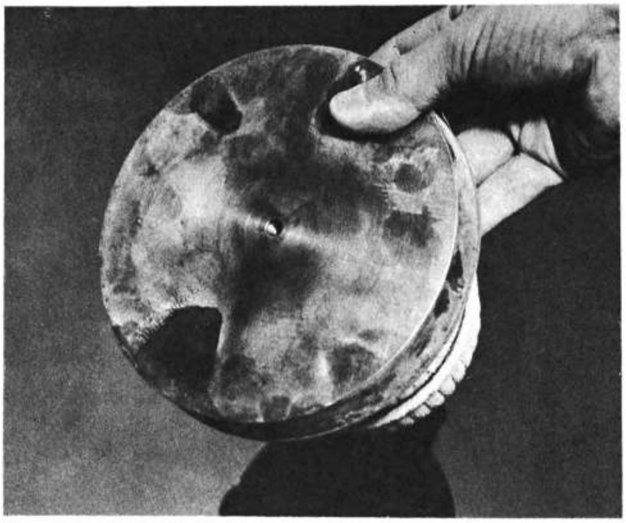
13–4 Galileo observed that a ball tends to rise to its original height regardless of the slope of the incline. With zero slope, the original height can never be reached. Therefore, motion along a horizontal plane should be perpetual.



13-5a The motion of a Dry Ice puck. It moved from left to right while the light was flashed 24 times in 10 sec, or every 10/24 sec. The scale at the bottom is in centimeters. Here is a close approach to the ideal situation of motion without force. The puck has a nearly constant displacement in equal time intervals.



13-5b A Dry Ice puck, as described in the text, rests on a smooth glass surface. With such apparatus we can study almost frictionless motion.



13-5c Dry Ice inside the puck changes into a gas which escapes through the hole in the bottom of the puck.

chanics. In the same way, we must look hard at very simple and idealized motions to obtain a real understanding of the basis of dynamics. Then, and only then, will we be ready to apply dynamics to the ordinary complex world.

With modern equipment we can do experiments that almost realize Galileo's idealized experiment on motion without force. To get the flash photograph of Fig. 13 – 5 (a), we used a Dry Ice puck sliding across a glass surface on a layer of gas. The gas bearing almost eliminates friction. In Fig. 13 – 5 (b) you can see the puck we used. It is a heavy metal disc, highly polished on the bottom surface and carrying a container full of pieces of Dry Ice. This frozen carbon dioxide changes slowly into a gas which escapes through a small hole in the center of the bottom of the metal disc [Fig. 13 – 5 (c)]. Thus a thin layer of gas is provided continuously between the metal disc and the glass surface; the puck floats on the gas layer.

The photograph of Fig. 13 – 5 (a) was taken by flashing a light 24 times in 10 seconds, or every $\frac{10}{24}$ sec, as the puck coasted over the smooth glass surface. As you can see, the displacement of the puck between flashes is almost constant. The velocity changes hardly at all. From such experiments we can calculate that if the puck were set going at about 10 kilometers per hour on a long enough horizontal surface, it would coast for about $\frac{1}{2}$ kilometer!

13-3 CHANGES IN VELOCITY WHEN A CONSTANT FORCE ACTS

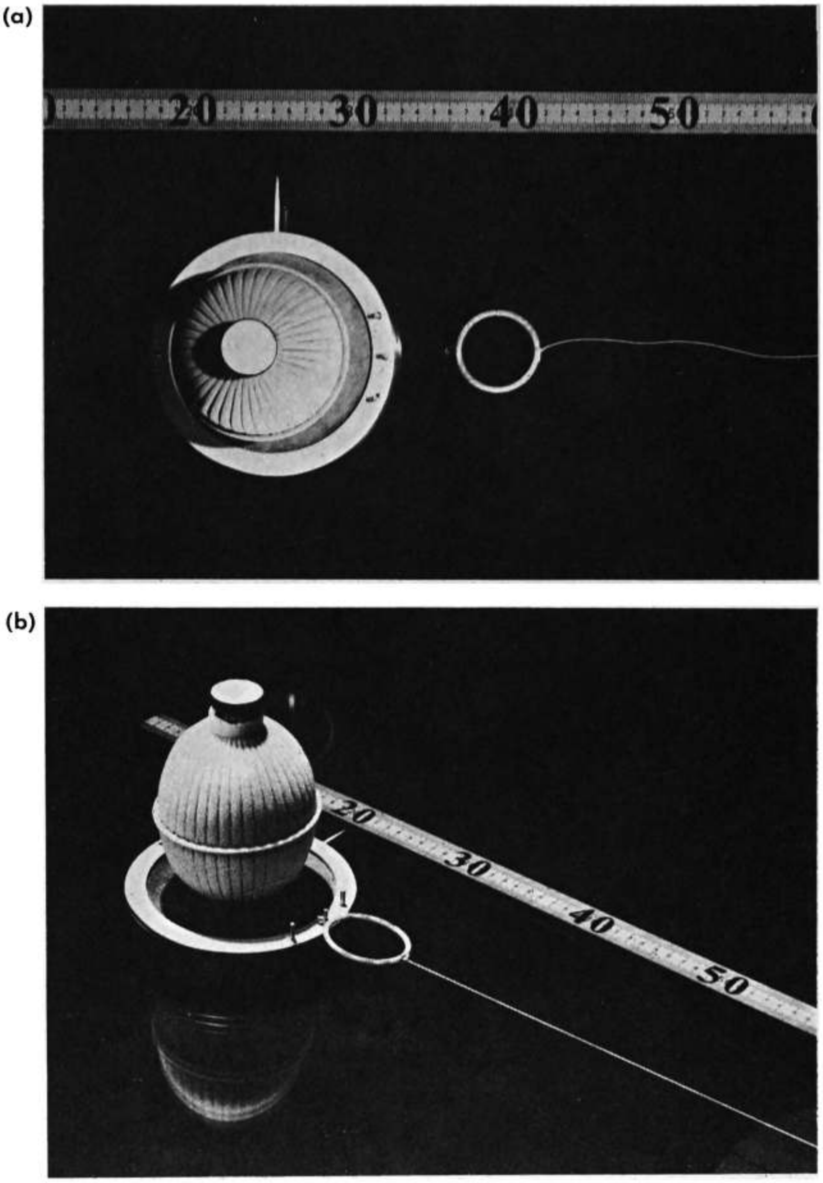
Galileo's law of inertia tells us that an object on which no force acts moves with unchanging velocity. If the velocity changes, we conclude that some force is acting on the object. What is the relation between force and change in velocity?

We shall begin our study of this question with the simplest experiment we can imagine. We shall apply a single force to a single object. To minimize other effects we use one of the same Dry Ice pucks on the same glass surface with which we examined motion with no force. The force we now apply is therefore the only force we need consider.

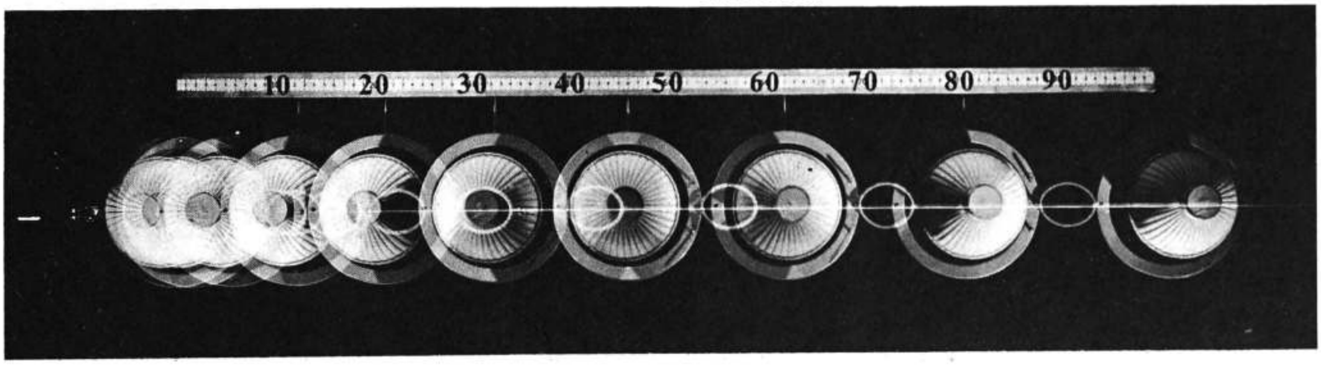
We need some way to recognize when the applied force is constant. For this purpose we employ a spring in the form of a loop attached to the puck [Fig. 13 – 6 (a)]. It is common experience that the force exerted by a spring increases in some manner with the stretch of the spring. Also, whenever a particular spring is stretched a definite amount it seems to exert the same force. We shall assume that the force exerted by our loop is the same whenever it is stretched the same amount [Fig. 13 – 6 (b)].

Now we pull our puck in such a way that the loop is always stretched the same amount, and record the motion of the puck by flash photography. The result of such an experiment is shown in Fig. 13 – 7. In this experiment the light flashed at intervals of $\frac{10}{24}$ second. We can see the successive positions of the puck, and we can see that the extension of the loop remained constant. The puck started from rest and moved on a straight-line path in the direction of the force we applied. Therefore, we can analyze this motion with the methods that we learned in Chapter 11.

Let us call the direction of motion of the puck the positive direction and label it as an x axis.



13–6 (a) Unextended spring loop before being mounted on puck. (b) The loop extended. Whenever this loop is stretched the same definite amount, we seem to get the same force.



13–7 The flash photograph shows the puck being pulled to the right. The light flashes were separated by 10/24 sec. A constant force was applied by keeping the loop extension constant. The displacement of the puck in each interval marked on the photograph has been measured and appears in Table 1.

Clearly, in each time interval the displacement, Δx, of the puck increased. Therefore, the average velocity $v = \frac{\Delta x}{\Delta t}$ increased (to the right) in each time interval. By measuring the successive displacements, we can establish quantitatively how the average velocity changed. As the caption under Table 1 shows, in each $\frac{10}{24}$ second after the motion started, the average velocity increased by about 5.8 cm/sec, a constant amount within the limits of accuracy of our experiment. Dividing the change of 5.8 cm/sec by the time interval of $\frac{10}{24}$ sec, we see that the average velocity changed at the constant rate of 14 cm/sec^2. Since the average velocity changed at a constant rate throughout the motion, we can safely assume that the instantaneous velocity also changed at the same constant rate. Therefore, the change of v in any interval of time Δt is

$$\Delta v = \left(\frac{14\text{ cm}}{\text{sec}^2}\right) \Delta t, \text{ to the right}$$

where the velocities are measured in cm/sec and the times are measured in seconds.

The particular value 14 cm/sec^2 occurs in this experiment because we pulled with a particular force on a particular object. When we pull with other forces or pull on other objects, we usually obtain other values of the proportionality factor. But all experiments like the one just described show that under influence of a constant force the velocity changes in direct proportion to the time the force acts.

TABLE 1 *Data from Experiment Shown in Fig. 13—7*

INTERVAL NO.	POSITION x(cm)	AVERAGE VELOCITY IN INTERVAL $\Delta x/\Delta t =$ v(cm/flash)	CHANGE IN AVERAGE VELOCITY Δv(cm/flash)
1	4.1	4.1	2.2
2	10.4	6.3	2.5
3	19.2	8.8	2.4
4	30.4	11.2	2.4
5	44.0	13.6	2.5
6	60.1	16.1	2.4
7	78.6	18.5	

The zero of position was taken as the location of the puck at the instant of the first flash. The value of x is the position of the puck at the end of the given interval. The average velocity, measured in cm/flash, is numerically equal to the displacement during the given interval. The last column simply gives the amount by which the average velocity changed between successive intervals. We find that Δv was constant within the limits of accuracy of our experiment and was 2.4 cm/flash. Since we had 24 flashes in 10 seconds, or 2.4 flashes/sec, the change in average velocity was (2.4 cm/flash) (2.4 flashes/sec) = 5.8 cm/sec between successive time intervals of duration 10/24 sec.

13–4 DEPENDENCE OF CHANGE OF VELOCITY ON MAGNITUDE OF FORCE

What happens when we apply a different constant force to the same body? Let us apply twice the force and see what happens. This simple suggestion raises a new problem. We have seen that we could use a spring loop to remove the human element from the operation of applying a force of a particular size, giving us confidence that we applied the same force throughout our last experiment. But how can we use the loop to apply twice the force?

One simple way of getting twice the force is suggested by the familiar fact that two men can push harder than one. For example, two men may be needed to push a stalled car which one alone could not move. We can arrange two

TABLE 2 *Data from Experiment Shown in Fig. 13—9*

INTERVAL NO.	POSITION x(cm)	AVERAGE VELOCITY IN INTERVAL $\Delta x/\Delta t =$ v(cm/flash)	CHANGE IN AVERAGE VELOCITY Δv(cm/flash)
1	8.4	8.4	4.7
2	21.5	13.1	4.7
3	39.3	17.8	4.8
4	61.9	22.6	4.8
5	89.3	27.4	

These are the results of an experiment in which the applied force was twice that used in the first experiment (Table 1). The flash rate was again 2.4 flashes per second. Note that the change of average velocity and hence of instantaneous velocity is just twice as great as it was before.

loops to give twice the force that one gives. Let us construct a second loop as nearly identical with the first as we can make it. When we stretch the second loop the same amount as the first, it should exert an equal force. We can assure ourselves that the forces are equal by doing our last experiment over again with the new loop. When the force is the same as before, the puck speeds up in just the same way as before. We thus show that the new loop gives the same force as the old one.

Now we can apply twice the old force to the puck. We hook both loops to the puck side by side and pull each of them in the same direction (Fig. 13 – 8). We make sure that each loop is extended the same amount as in the original experiment with one loop, and we observe the motion in the same way as before. In this way we can double the force on the puck, leaving everything else the same.

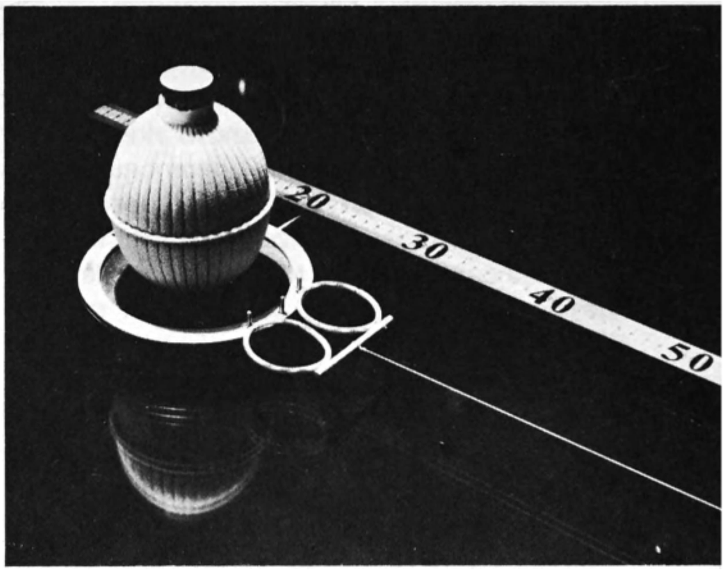
We have used this procedure to put a doubled force on the same puck as before. The result is shown in Fig. 13 – 9. What do we find? The data in Table 2, taken with exactly the same arrangement as before but with the force doubled, shows that the velocity increases twice as fast as before. Instead of $\Delta v = (14\ \text{cm/sec}^2)\Delta t$ we obtain $\Delta v = (28\ \text{cm/sec}^2)\Delta t$, where again we measure the velocities in cm/sec and the times in seconds.

Further experiments show that this result is general. Whenever we double the force on a given object, we double the rate at which its velocity changes. Moreover, if we make the force three times as large by putting three identical loops side by side, we triple the rate at which the velocity changes. From many measurements of this kind we conclude that the change in velocity of a body in a given time interval is proportional to the force F that acts on the body.

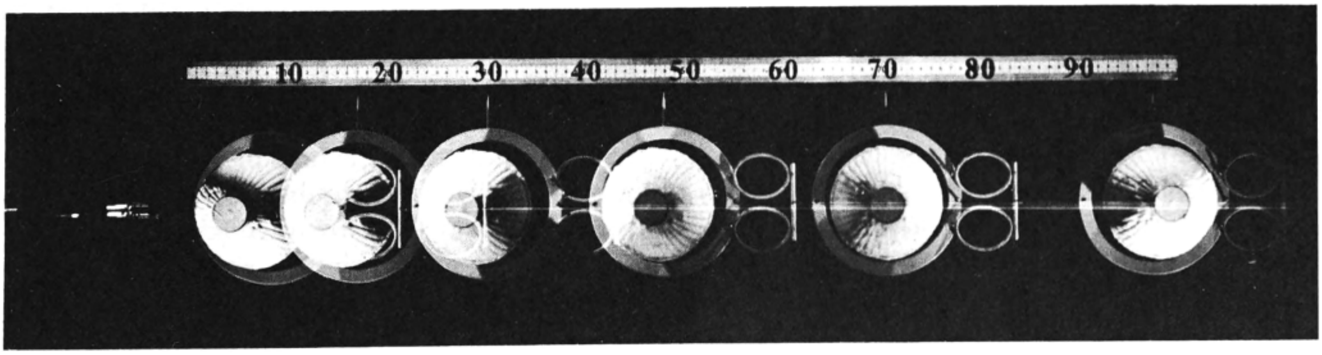
We have now found two things. The change in velocity Δv increases with the length of time interval Δt, and it is greater, the greater the force. We can combine this information in one statement: Δv is proportional to $F\Delta t$.

13–5 INERTIAL MASS

The change Δv produced by a given force F acting for a given time Δt depends on the object on which the force acts. Applying equal forces for the same time to a tennis ball and to an elephant produces less change in the velocity of the elephant than in that of the tennis ball.



13–8 To apply twice the original force, we attach two identical loops to the puck.



13–9 The puck used in the first experiment is accelerated by twice the force used before. The interval between flashes is again 10/24 second.

Because larger bodies are less easily accelerated by forces, it is convenient to write the proportionality between $F\Delta t$ and Δv in the form:

$$F\Delta t = m\Delta v$$

The proportionality constant m depends on the object. Its value increases with the size of the body, at least for objects made of one uniform substance.

The constant m is called the *inertial mass* of the body. By rearranging the above equation, we define m as $F\Delta t/\Delta v$ for a given object. This ratio is experimentally constant and it tells how difficult it is to change the velocity of the body. As you already know, $\Delta v/\Delta t$ is the rate of change of velocity, or the acceleration a, in the direction of the force. The inertial mass can therefore be written $m = F/a$. The greater the force needed to produce a given acceleration, the greater the inertial mass of the object.

Naturally we wish to know whether inertial mass, this measure of the difficulty of accelerating a body, is a new property of a body. Is it unrelated to anything we know already, is it some familiar property in a different guise, or is it some combination of familiar properties? To answer these questions, we investigate the relation of the inertial mass to the shape, size, composition—in fact to any other known property of an object.

Of course, we do not look about at random. As the example of the elephant and tennis ball shows, our experience already indicates that size is a reasonable place to start. But volume alone will not do. A hollow elephant, on a float in the Mardi Gras parade, is not so hard to accelerate as a real elephant; so let us start with objects of uniform composition. Let us try the effect of doubling the size of an object while keeping its composition uniform. One easy method is to use two identical pucks. We can decide whether two pucks are identical by subjecting them to the same force for equal times. If each gains the same speed, they have identical inertial masses.

Let us pull the two identical pucks side by side, applying the same force to each with identical loops. The two pucks travel along together, each gaining the same Δv as they progress. Next we join the pucks rigidly together and pull them with both loops yoked side by side. Again, we would expect them to gain speed in the same way. Experiments show that this is so. Now let us try pulling the connected pucks with the force of only one loop. This is half the previous force; and from what we know already, we expect that the double puck will gain speed at only half the previous rate. Again experiment confirms our expectation. In other words, when we apply one standard force to two standard pucks, the doubled puck speeds up just half as fast as a single puck does.

In all experiments with a double puck the ratio F/a is doubled. But this ratio is defined as the inertial mass of the body; so we have doubled the inertial mass by using two identical pucks. These experiments indicate that mass is directly proportional to the size of bodies made from the same material. Further experiments, using various numbers of pieces of the same substance, all agree with this conclusion.

There is nothing magic about pucks and gas bearings. We can investigate the dependence of inertial mass on the number of pieces by joining together numbers of identical well-loaded model railroad cars, with carefully made bearings so that there is little friction in the wheels. Such experiments are not quite so close to the ideal of eliminating all forces except those we apply on purpose, but the proportionality of the inertial mass to the number of cars is still clearly evident.

Our results for identical bodies are not astonishing. The real questions about inertial mass arise when we consider objects made of different materials. We cannot have a piece of silver identical with a piece of gold. But we can find pieces that have the same inertial mass, pieces for which F/a is the same. Such pieces are certainly not identical in size or composition. Inertial mass is, then, not entirely a matter of size.

What happens when we join together a piece of silver and a piece of gold of the same inertial mass? To find the inertial mass of the new body we pull it with various forces F for certain times Δt. We find that F/a, the inertial mass of the new body, is just twice the inertial mass of each original piece. Thus when we add the pieces together, we also add their inertial masses. In fact, we can take any piece of gold and any piece of silver and measure their individual inertial masses, m_1 and m_2. When we join these bodies together we find that the inertial mass, m, of the combination, again measured as F/a, is now equal to the sum of the two original inertial masses: $m = m_1 + m_2$. The same is true for bodies made of any other substances. Inertial masses are additive.

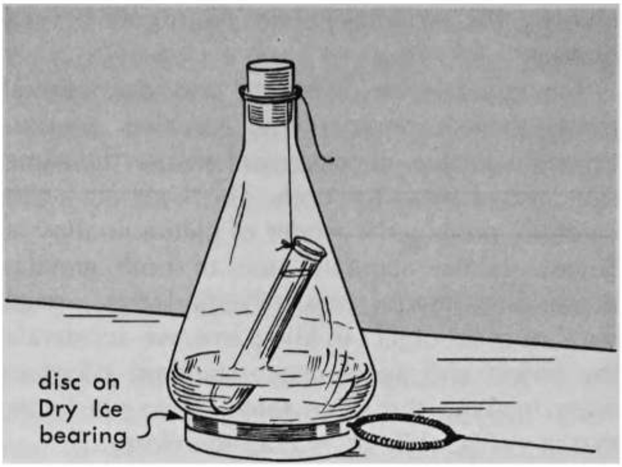
In concluding that inertial masses add, we have concluded that the inertial mass does not depend on the shape or the chemical nature of an object. We did not specify how we combined silver and gold pieces, nor do we need to. If we melt down any object and cast it again in any shape we choose, the ratio F/a will not be affected.

Let us go even further. Suppose we measure the inertial mass of a flashbulb, then flash it so that the magnesium and oxygen inside combine to form magnesium oxide. We find that the chemical combination has not affected the inertial mass. Or, we can place separate solutions of sodium carbonate (washing soda) and calcium chloride in a closed vessel (Fig. 13 – 10). We measure the inertial mass of this system, then upset the container so that the substances react to form calcium carbonate (an insoluble white solid) and a solution of table salt. When we measure the inertial mass again, we find no change.

What, then, do we know about inertial mass? It increases in proportion to the amount of a single substance in an object. When various pieces of matter are put together, the inertial masses add irrespective of the nature of the materials involved. Finally, inertial mass is conserved in chemical reactions.

13–6 INERTIAL AND GRAVITATIONAL MASS

The properties of inertial mass remind us of the properties of mass as measured on a balance (Chapter 1). When a balance is in equilibrium we say that we have equal masses on each side. The masses thus measured are called *gravitational* masses because at equilibrium the gravitational pull of the earth on each mass is the same. Actually the earth is not important; the property we measure is a property of the body alone. The balance works just as well at the top of a mountain, where the pull of the earth on each object is weaker. It would work as well on the moon, where the masses we compare would be pulled still more weakly. The only important thing in measuring gravitational mass is that we compare gravitational pulls on the objects when they are in the same place in relation to the other pieces of matter in the universe.



13–10 Apparatus used in an experiment to show that inertial mass does not change in chemical reactions.

In terms of the measurements by which we determine them the gravitational mass and the inertial mass have no connection. To measure inertial mass we apply a force to an object and find its acceleration. Gravity is irrelevant. On the other hand, when we measure gravitational mass by using a balance in equilibrium under gravitational forces, we do not have any motion and we must have gravitational forces. Two measurements could hardly differ more completely. Nevertheless the properties of the gravitational mass are strikingly similar to those we have just found for inertial mass. The gravitational mass of one substance is proportional to the amount of the substance present. Gravitational masses of any substances add. Gravitational mass is conserved in chemical reactions.

The additivity of masses of each kind and the conservation of each kind of mass in chemical reactions suggest that the gravitational and inertial mass may be proportional for any object. This proportionality can be tested by measuring the inertial and gravitational masses of many different objects of many different compositions. In fact, such experiments have been done many times over. To within our best experimental accuracy, the inertial masses of all objects are proportional to their gravitational masses.

The equivalence of inertial and gravitational mass—their experimentally revealed proportionality—makes it convenient to use the same standard of mass for both. The kilogram—that carefully protected cylinder of platinum alloy at Sèvres—is the standard unit of both gravitational and inertial mass. To find the inertial mass m of an object in kilograms, we accelerate the object and accelerate a standard kilogram mass m_s with the same force. Then we know that $m = F/a$ while $m_s = F/a_s$; therefore,

$$\frac{m}{m_s} = \frac{a_s}{a}$$

Because $m_s = 1$ kg, the mass m in kg is given by the ratio a_s/a. For example, if a certain force accelerates a kilogram mass at $\frac{1}{2}$ m/sec² and another object at 2 m/sec², the mass of the other object is $\frac{1}{4}$ kg.

It is often difficult to establish the idealized conditions in which we can attach a particular force to an object and can be sure that no unknown forces influence the resultant motion. But we need not measure the inertial mass of every object in which we are interested in this direct way. Because of the apparent equivalence of inertial and gravitational mass, a measurement of gravitational mass tells us the inertial mass. We need not normally worry about the distinction between them, and we usually use the word "mass" alone to refer to either one.

13–7 NEWTON'S LAW: DYNAMICAL MEASUREMENT OF FORCE; UNITS

The relation $F\Delta t = m\Delta v$ tells us how the change in velocity Δv is related to the inertial mass m, to the cause of motion F, and to the time Δt that the force acts. This relation embodies Newton's law of motion.*

When we know the mass of an object, we can use Newton's law in either of two ways. We can predict the change in velocity of the object if we know the force F and the time Δt that it acts; or we can determine the force by observing the change in velocity it imparts to the object. When we observe the acceleration of a body and thus determine the force acting on it, it is convenient to write Newton's law as

$$F = m\frac{\Delta v}{\Delta t}$$

$$\text{or}\quad F = ma$$

showing explicitly that the force is proportional to the acceleration. In this form the cause of the motion is on one side of the equation and the inertial mass of the body and its motion are on the other.

Suppose that we observe the same object in two different experiments. Perhaps we see that it speeds up three times as fast in the second experiment as in the first. We can then conclude that the force acting in the second experiment is three times as great as the force in the first experiment. In other words, we can use the acceleration $a = \Delta v/\Delta t$ of a given object as a measure of the force, just as we used it earlier to establish the equality of the forces exerted by two identical spring loops when they produced equal acceleration on the same object.

We now have a dynamical method to determine both inertial masses and forces. We start with a kilogram mass. We use the force which accelerates it at the rate of 1 m/sec² as our unit of force. It is called a *newton*. To measure another mass we determine its acceleration when acted on by our 1-newton force. By definition its inertial mass is F/a; and in this case where the force is 1 newton, the mass in kilograms is $m = (1\text{ newton})/a$, where a is expressed in meters/sec². In this way we have used one standard mass to give us a standard force. Then we have used the standard force and the relation $m = F/a$ to measure the masses of other objects that interest us.

We can also use the same relation to find other forces. From $F = ma$ we can determine any force which acts on one of our known masses by measuring the acceleration it produces. The acceleration in meters/sec² times the mass in kg gives the force in newtons.

---

*What we call Newton's law is often called his second law, and Galileo's principle of inertia, which is a special case of the general law, is sometimes called Newton's first law. The names of Newton's first and second laws don't change the content, but it may be important to know them so that you can understand what someone means when he says, "According to Newton's first law. . . ."

13–8 FORCES THAT CHANGE, AND NEWTON'S LAW

We have discussed Newton's law of motion under what appear to be rather special conditions. For simplicity, we have accelerated masses from rest using a constant force. Does the connection that we have found between force and acceleration hold equally well if we change the magnitude of the force while a body is in motion?

Suppose we push some object, initially at rest, with a steady force for a definite time. It will gain speed as long as we push it. If we stop pushing, acceleration ceases; the body moves on at constant velocity. If we begin pushing again, we again get acceleration. Suppose we apply a force in the direction opposite to the motion. Then we expect the acceleration to be in the direction of the force. Because it is opposite to the motion, the object should slow down instead of speeding up. Experiments show that the rate of slowing down is indeed F/m, where F is the force and m the inertial mass.

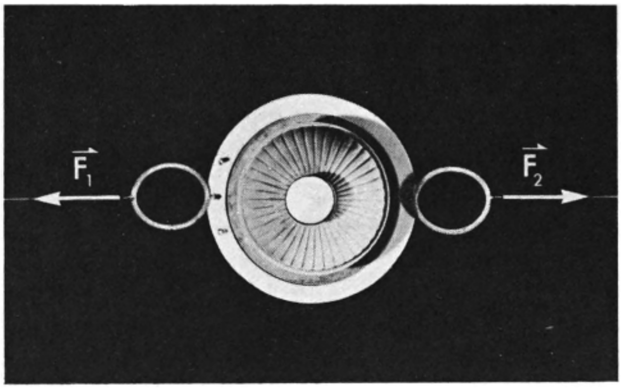
Whether a body is standing still or sailing through outer space at 10^5 m/sec, a force will accelerate it: we always find $a = F/m$. We need not consider what velocity the body now has, nor what process produced this velocity. No matter what the past history or the present motion of a body, a given force applied along the line of motion will produce the same acceleration. The equation $F\Delta t = m\Delta v$ emphasizes that we can chop out any period of time, large or small, and during this Δt we will find a definite Δv depending only on the inertial mass and the applied constant force—no matter what comes before or after in the progress of the body.

What we have said in this section is observed experimentally to extremely high accuracy. But when we observe masses moving at increasingly greater speeds, slight discrepancies between the observed behavior and Newton's law become apparent. By the time the speed has increased to 10^8 m/sec, deviations of a few percent occur. The equation $F\Delta t = m\Delta v$ is no longer an adequate description. By extending the range of observations, we thus learn that Newton's law must be modified. It must be extended so that in its new form it makes slightly different statements about objects moving at high enough speeds. On the other hand, at lower speeds it must continue to say the same thing that we have stated here. Einstein and others have formulated the needed extension for ultrahigh speeds. This modification does not overthrow Newton's law: it includes and extends it

13–9 HOW FORCES ADD; THE NET FORCE

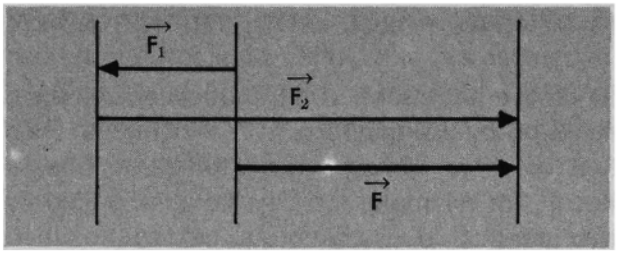
So far we have studied the motion of an object acted on by a single force. What happens when two or more forces act on the same object? Recall, for example, the Dry Ice puck pulled by two identical loops yoked one next to the other. As we learned, with this arrangement the force on the puck is twice that of the single loop and the acceleration that results is twice the acceleration imparted by one loop. The acceleration is proportional to the sum of the two individual forces.

We can also make two identical loops, each stretched the same amount, pull in opposite directions (Fig. 13 – 11) Then no acceleration takes place. For example, if you and a friend pull equally hard on a book, but in opposite directions, the book does not accelerate. Because they are oppositely directed, the forces on the book add to zero. Apparently the net force, the force which changes the motion, is obtained by adding the forces in the same way that we added displacements or vectors. As far as their effect on the motion is concerned, two forces of equal magnitude and opposite direction just cancel, and one of them can be considered to be the negative of the other.

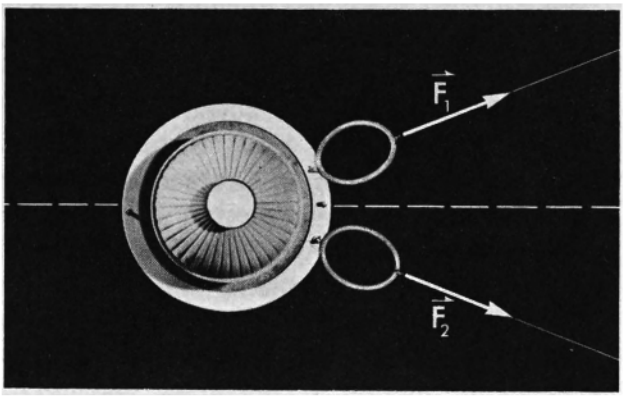


13–11 Two forces, equal in magnitude but opposite in direction, are acting on a Dry Ice puck. The net force is zero, and the observed acceleration is zero.

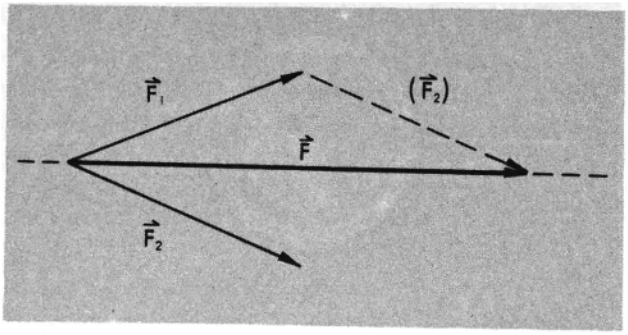
In general, when we apply forces on an object in opposite directions, the acceleration of the object is found to be proportional to the sum of the forces taken with the direction included. When a force of 1 newton acts to the left, and a force of 3 newtons pulls to the right on a Dry Ice puck, the puck accelerates to the right in just the way it does when a single force of 2 newtons acts on it. The net force of 2 newtons is the sum of the individual forces taken as indicated in Fig. 13–12 Furthermore, when any number of individual forces act on an object, we find that Newton's law of motion holds and the observed acceleration arises from the net force.



13–12 A force $\vec{F}_2$ of 3 newtons is acting to the right, and a force $\vec{F}_1$ of 1 newton is acting to the left. Adding them like displacements, we obtain the sum $\vec{F}$ of 2 newtons acting to the right. This is the net force that appears in Newton's law.



13–13 Two forces of equal magnitude act at an angle to each other. The object accelerates along the dotted line bisecting the angle between the directions of the two forces. We conclude that the net force acted along this line.



13–14 The vector addition of the two forces that are shown in Fig. 13–13. The sum is the net force which determines both direction and magnitude of the acceleration of a given mass.

Two forces need not pull in the same or in opposite directions; they may be pulling at an angle to each other. What is then the direction and magnitude of the net force? Suppose we pull equally hard on an object with each of two identical loops, as in Fig. 13–13. We find that the object accelerates along the line bisecting the angle between the directions of the two forces, the line dotted in the figure. Apparently we have applied a net force along the dotted line. It is easy to jump to the conclusion that the net force is the vector sum of the two separate forces. Experiments show that this is indeed the case. The acceleration imparted by the two loops in Fig. 13–13 is given by $F = ma$, where F is the magnitude of the vector that we obtain by treating the individual forces as vectors and adding them together to get the net force vector (Fig. 13–14). Also, when the two forces are not equal or when there are more than two forces, the magnitude and direction of the net force is given by the vector sum of the individual forces. This net force determines the acceleration in accord with $a = F/m$.

Let us sum up what we have learned so far. We started out by studying the acceleration of bodies from rest under the influence of a single force. This led us to Newton's law of motion. We then investigated what happens to a body which is already moving when a force acts on it in the direction of motion or opposite to it. We found that Newton's law still holds. We then asked what happens if several forces act on a body. Again, Newton's law holds as if a single force, the net force, acts on the body. The net force is the vector sum of all the forces.

13–10 THE VECTOR NATURE OF NEWTON'S LAW

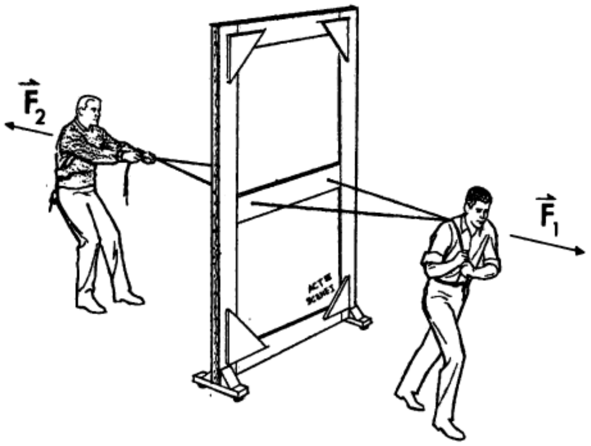
Newton's law is even more general than we have yet indicated. So far we have applied a net force only along the direction of motion of an object or opposite to it. As a result of these net forces we change the speed of the object but not the direction of its motion. Forces, however, can be applied in other directions. We can shove a moving ball out of its course by pushing across its direction of motion. In general, then, we see that forces change the velocity vector describing the motion of an object in either magnitude or direction, or both. The force is itself a vector, and Newton's law connects it with the rate of change of the vector velocity. We should write the law as

$$\vec{\mathbf{F}}\,\Delta t = m\Delta\vec{\mathbf{v}}$$

or

$$\vec{\mathbf{F}} = m\frac{\Delta\vec{\mathbf{v}}}{\Delta t} = m\vec{\mathbf{a}}$$

where $\vec{\mathbf{a}}$ is the vector acceleration. In the next chapter we shall see some of the evidence for the vector nature of Newton's law and discuss some of its implications.



13–15 The force we exert may not be the only force that is acting.

13–11 FORCES IN NATURE

When we pull an object, we cannot usually be sure that the force we exert is the only force acting on the object (Fig. 13 – 15).

Sometimes the origin of forces on an object is not immediately apparent, or forces may arise as a result of the motion of the object. For example, the wind pushing on the surface of a balloon can exert a force, and we must pull on the balloon in the opposite direction to prevent it from blowing away. Even if the air is still, when we move the balloon through it, the air gives rise to a force opposing the motion. When the balloon is in motion, its acceleration is not given by the force we apply. It is given by the net force. If the net force is zero, the acceleration is zero and the balloon moves at constant velocity.

We can measure the force exerted by a steady wind going past the balloon by measuring the extension of a loop in a string holding the balloon still. This force increases with the speed of the wind. As we all know, when the wind blows faster, it blows harder. When we then pull the balloon through still air, we must subtract this force from the force we exert to get the net force that is responsible for the acceleration.

Frictional forces become evident when we try to move an object along a surface. Unlike the retarding force which arises when we pull a balloon through the air, frictional forces are often nearly independent of the speed of the object. It was undoubtedly the common occurrence of friction that led the Greeks to conclude that force was needed to maintain a constant motion. The force needed is equal and opposite to the forces of friction and air resistance. The net force is zero.

The idea of force as a cause of motion is valuable because it enables us to predict what motion will occur in a given situation. The same forces occur over and over again whenever the same situation arises. Some forces are independent of motion; for example, the force of gravitational attraction, the weight of a body. As we shall see in the next chapter, this force is the same, whether the body moves or stands still. If we know our geographical position, we know what gravitational force to expect, and we can make predictions about the motions of a falling object. Other forces depend on the relative motion of one body with respect to another. One of the essential jobs we face is to learn about the forces in nature. Then we can use the observed forces to make predictions about motion and to design mechanical apparatus.

QUESTIONS

1.* What is the difference between *kinematics* and *dynamics?*

2.* Was it necessary for Galileo to measure time in order to get the extrapolation which led him to the law of inertia?

3. A ball is released from rest on the left-hand incline of Fig. 13 – 4 at a height of 10 cm above the lowest point.

(a) If there is no friction, how high vertically will it rise on the right-hand incline?

(b) If the right-hand incline rises 1 cm for every 10 cm of horizontal distance, how far will the ball travel horizontally on it?

(c) If the incline rises only $\frac{1}{2}$ cm for every 10 cm of horizontal distance, how far will the ball go?

4.* If you roll a putty ball around the inside of a bowl, it soon comes to rest in the center. A steel ball will orbit quite a few times before coming to rest in the center. What would be the "ideal" motion of a ball in a bowl?

5. Why is it particularly dangerous to drive on an icy highway?

6.* If a ball is rolling with a speed of 20 cm/sec and no force acts on it, what will be its speed after 5 sec?

7.* In Table 1, what would you predict for the average velocity during the eighth time interval if the puck continued to be pulled by the stretched loop?

8.* Suppose a Dry Ice puck increases its velocity so that it travels first at 10 cm/sec, at 12 cm/sec at the end of the next second, 14 cm/sec at the end of the second after that, and so on. What can you tell about the force acting on the puck?

9. A body is pulled across a smooth horizontal surface by a spring loop that is kept stretched by a constant amount. It is found that the body is accelerated at 15 cm/sec². What will be the acceleration of the body if it is pulled by two loops, each just like the first loop, side by side and stretched by that same amount? (See Fig. 13 – 8.)

10. An object sliding on a low-friction bearing is pulled with a constant force. In a time interval of 0.3 sec the speed changes from 0.2 m/sec to 0.4 m/sec. In a second trial, the object is pulled with another force. In the same length of time the speed now changes from 0.5 m/sec to 0.8 m/sec.

(a) What is the ratio of the second force to the first?

(b) If the body is pulled with the second force for 0.9 sec, what change in speed results?

(Note that the forces are in the direction of motion.)

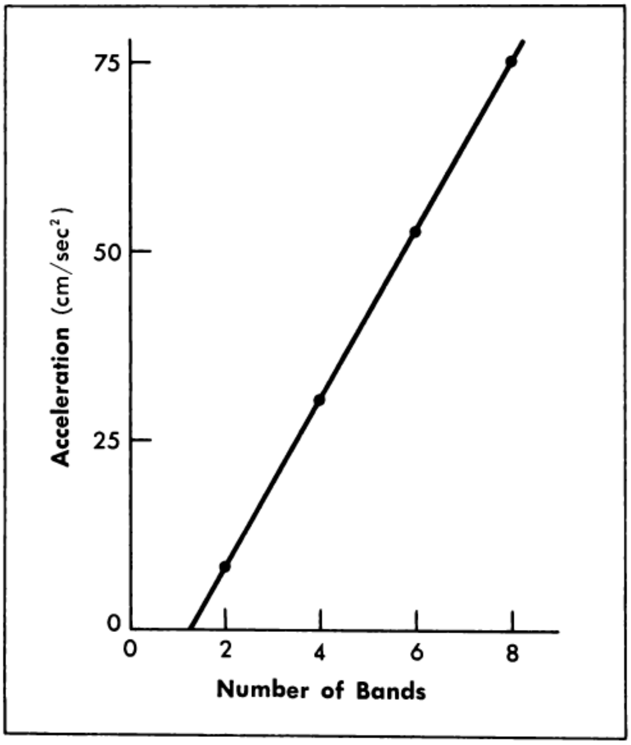
11. A block is pulled along a horizontal surface by 2, 4, 6, and 8 parallel bands of rubber. All bands are alike and each is stretched the same length in each experiment. The graph (Fig. 13 – 16) shows the resulting accelerations plotted versus the number of bands.

(a) What can you conclude from the fact that the plot yields a straight line?

(b) What does the intercept of the graph with the horizontal axis measure?

(c) Can you use the graph to predict the acceleration of the block produced by stretching one rubber band by the standard amount?

(d) Suppose you repeat the experiment, changing only the surface on which the block is pulled. How will the new graph relate to the old one?



13–16 For Question 11.

12.* In Fig. 13 – 7, suppose the photograph represented a body moving from right to left. In which direction would the force be acting in this case?

13.* A car can accelerate by 3 m/sec². What is its acceleration if it is towing another car like itself?

14. Why is the flask in Fig. 13–10 closed? Be prepared to explain in class.

15. Why do we make such careful distinctions between gravitational and inertial mass, rather than talking about one mass only, since they are equivalent?

16.* Two objects are accelerated separately by the same force. Object A accelerates at 20 cm/sec² and B at 60 cm/sec². What is the ratio of (a) their inertial masses, (b) their gravitational masses?

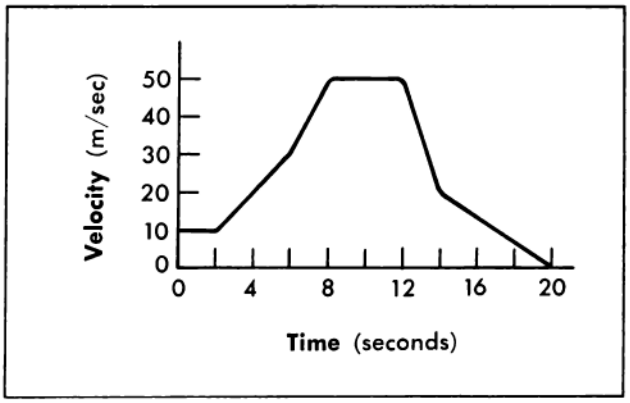
17.* A carton is balanced by two identical cans on a balance. A certain force accelerates the carton by 2 m/sec². How much will the same force accelerate one of the cans?

18.* A body with a mass of 0.5 kg is accelerated at 4 m/sec². How large a force is acting?

19.* A boy gives a 0.5-kg ball a speed of 6 m/sec from rest in 0.2 sec. What average force did he apply to the ball?

20. A force of 5 newtons gives a mass m_1 an acceleration of 8 m/sec², and a mass m_2 an acceleration of 24 m/sec². What acceleration would it give the two when they are fastened together?

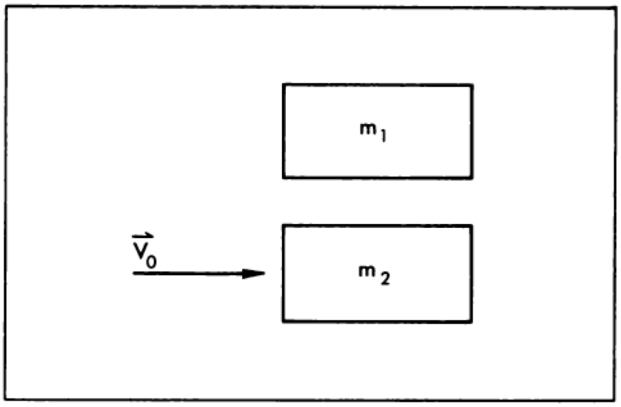
21. The graph in Fig. 13 – 17 shows the velocity, along a straight line, of an object of mass 2 kg, as a function of time. Plot a graph of the force as a function of time.



13–17 For Question 21.

22. A block of mass 3.0 kg is moving along a smooth horizontal surface with a velocity v_0 at an instant of time $t = 0$. A force of 18 newtons is applied to this body opposite to the direction of its motion. The force slows the block down to half its original velocity while it moves 9.0 m.
   (a) How long does it take for this to occur?
   (b) What is v_0?

23. Consider the two masses m_1 and m_2 in Fig. 13 – 18. At time $t = 0$, m_1 is at rest and m_2 is moving with velocity v_0. Is it possible to apply the same constant force (in magnitude and direction) for the same time to both masses to bring them to equal velocity at a given moment? Try to solve this problem by a qualitative argument and then check your conclusion by writing down the necessary equations. Consider all cases: $m_1 < m_2$, $m_1 = m_2$, $m_1 > m_2$.



13–18 For Question 23.

24. You observe an object covering distance in direct proportion to t^3, where t is the time elapsed.
   (a) What conclusion might you draw about the acceleration? Is it constant? Increasing? Decreasing? Zero?
   (b) What might you conclude about the forces? Be prepared to discuss in class.

25.* If several forces of different magnitudes and directions act on an object, in what direction will the object accelerate?

26.* In Fig. 13 – 11, what would be the acceleration of the puck if its mass were 0.50 kg, $F_2 = 10$ nt, and $F_1 = 6$ nt?

27.* In Fig. 13 – 13, suppose $F_1 = F_2 = 2.0$ nt. What would be the net force on the puck?

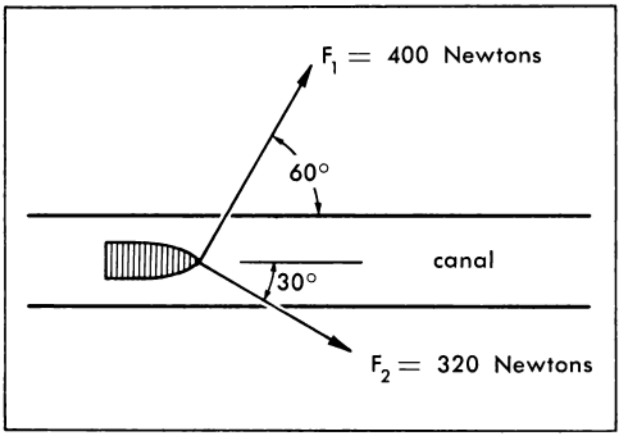
28. A block of mass 8.0 kg, starting from rest, is pulled along a horizontal tabletop by a constant force of 2.0 newtons. It is found that this body moves a distance of 3.0 m in 6.0 sec.
   (a) What is the acceleration of the body?

(b) What is the ratio of the applied force to the mass?

(c) Since your answer to part (b) is not equal to that to part (a) (at least, it shouldn't be), what conclusions can you draw about this motion? Give numerical results, if possible.

29. Two men wish to pull down a tree by means of a rope fastened near the top. If they use only one rope, the tree will come down on top of them. To prevent this, they tie two ropes 10.0 meters long to the same point and stand on the ground 10.0 meters apart when they pull. If each pulls with a force of 300 newtons, what is the force exerted by the ropes on the tree?

30. Two men and a boy are pulling a boat along a canal. The two men pull with forces $\vec{\mathbf{F}}_1$ and $\vec{\mathbf{F}}_2$ whose magnitudes and directions are indicated in Fig. 13 – 19. Find the magnitude and direction of the smallest force which the boy could exert to keep the boat in the middle of the canal.

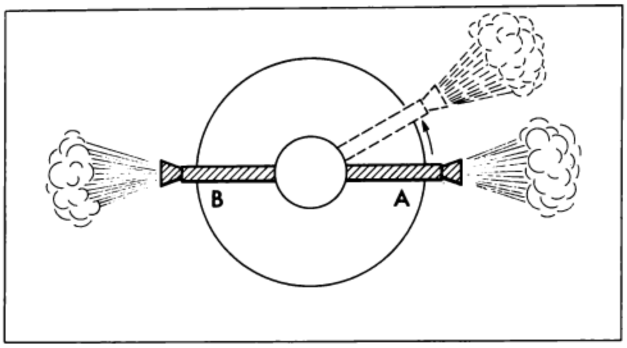


13–19 For Question 30.

31. Two small identical rockets *A* and *B* (Fig. 13 – 20) are mounted on a large Dry Ice puck of mass *m*, and the puck is placed on a flat surface. The rockets are ignited in the position shown and cause no motion. After the rockets are burning, a mechanism causes rocket *A* to rotate uniformly counterclockwise through one revolution while rocket *B* continues to point in the same direction.

(a) Draw a vector diagram showing the resultant acceleration at each one-eighth of a period.

(b) From the vector diagram you have drawn for part (a) determine the final velocity of the puck in terms of the acceleration produced by one rocket and the period of revolution *T* of rocket *A*.



13–20 For Question 31.

32.* You have to push with a force of 200 nt to slide a refrigerator across a floor at constant speed. What is the force of friction acting on the refrigerator?

33. The retarding force of air resistance on a balloon is proportional to the square of the velocity. For a certain balloon, inflated a certain amount, this force is given in newtons by $F_R = .2v^2$ where v is the velocity in m/sec. The balloon and the air inside have a combined mass of 10 gm.

(a) Draw graphs of the balloon's acceleration as a function of velocity when you pull it with a 1.8-newton force and with a 7.2-newton force.

(b) What is the maximum velocity that the balloon will reach in each case?

(c) If the mass were 5.0 gm, how would this affect the maximum velocity?

(d) What do you think would be the effect on the maximum velocity if you inflated the balloon to a larger volume?

34. Aristotle taught that a constant force was required to produce a constant velocity and from this he concluded that, in the absence of force, bodies would come to rest.

(a) Name several situations where a constant force seems to produce a constant velocity.

(b) How do you explain each of the situations in (a) in the light of Newton's law of motion?

35. How would you define a unit of mass if a standard spring were used in the metric system instead of a standard mass?

14

Motion at the Earth's Surface

IN THIS chapter we shall start our study of the forces occurring in nature. We shall determine them from the motions of the bodies on which they act, or, when possible, by balancing the unknown forces against known forces that we can apply. Here on the surface of the earth the force of gravity which pulls objects toward the center of the earth is familiar to us all. We shall first study this gravitational attraction.

14–1 WEIGHT AND THE GRAVITATIONAL FIELD OF THE EARTH

We know that objects on the surface of the earth are pulled toward its center. Different objects are pulled down by gravitational forces of different magnitude. The magnitude of this force on an object is called the weight of the object. It may be measured by hanging the object on a spring calibrated in newtons.

If we take a standard kilogram mass and measure its weight anywhere on the surface of the earth, we find it to be very nearly 9.8 newtons. Actually the weight of an object differs slightly from place to place on the earth's surface. However, the fractional variation in weight is the same for all objects. For example, at the North Pole a 1-kg mass weighs 9.83 newtons and at the equator it weighs 9.78 newtons. At the pole a mass of 2 kg weighs 19.66 newtons, and at the equator it weighs 19.56 newtons. Both masses change weight by $\frac{1}{2}$ percent, when moved from pole to equator. The weight of the 2-kg mass is always exactly twice that of a 1-kg mass at the same place. Therefore, the gravitational force $\vec{\mathbf{F}}$ acting on an object of gravitational mass m_g can be written

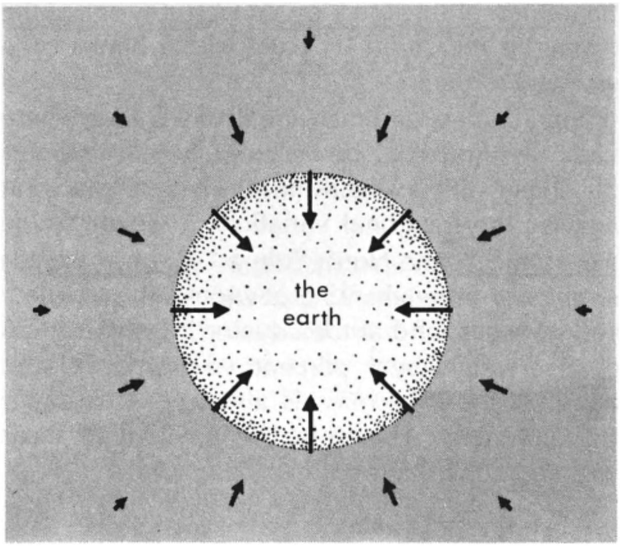
$$\vec{\mathbf{F}} = m_g\vec{\mathbf{g}}$$

where $\vec{\mathbf{g}}$ is the proportionality factor between the gravitational force and the gravitational mass. The mass is independent of position on the earth's surface—if two masses balance (on an equal-arm balance) at one place on the earth, they balance at all other places. Therefore, we learn that the proportionality factor $\vec{\mathbf{g}}$ varies slightly from place to place near the surface of the earth, and it is the same for all masses at any one place.

TABLE 1 *Gravitational Field of the Earth*

PLACE	LATITUDE	ALTITUDE, METERS	MAGNITUDE OF THE FIELD, NEWTONS/KG
North Pole	90°	0	9.832
Greenland	70°	20	9.825
Stockholm	59°	45	9.818
Brussels	51°	102	9.811
Banff	51°	1376	9.808
New York	41°	38	9.803
Chicago	42°	182	9.803
Denver	40°	1638	9.796
San Francisco	38°	114	9.800
Canal Zone	9°	6	9.782
Java	6° South	7	9.782
New Zealand	37° South	3	9.800

The factor $\vec{\mathbf{g}}$ is the gravitational force per unit mass. It is a vector with the dimensions of newtons per kilogram. We can measure and tabulate it for a large number of locations on the surface of the earth. Such a collection of physical quantities depending on position is called a field; and in particular the gravitational force per unit gravitational mass at different locations relative to the earth is called the gravitational field of the earth. Figure 14 – 1 represents part of the gravitational field around the earth, and Table 1 gives its magnitude at a number of different places.



14–1 The gravitational field of the earth. The inner vectors represent the magnitude and direction of the field at the earth's surface. The middle set gives values for a height of 3×10^6 meters. The outer vectors are for a height of 6×10^6 meters. (The radius of the earth is 6×10^6 meters.)

14–2 FREE FALL

How does an object move in the gravitational field $\vec{\mathbf{g}}$ of the earth? Suppose we let it fall. Its motion depends on the net force acting on it and on its *inertial* mass. The net force $\vec{\mathbf{F}}$ acting on the falling body is given by the vector sum of the gravitational attraction $\vec{\mathbf{F}}_g = m_g\vec{\mathbf{g}}$ (the weight of the body) and the air resistance $\vec{\mathbf{F}}_a$. It is

$$\vec{\mathbf{F}} = \vec{\mathbf{F}}_a + m_a\vec{\mathbf{g}}$$

According to Newton's law of motion, the acceleration of the object is proportional to this net force and inversely proportional to the inertial mass m_i; that is,

$$\vec{\mathbf{a}} = \frac{\vec{\mathbf{F}}}{m_i} = \frac{\vec{\mathbf{F}}_a + m_g\vec{\mathbf{g}}}{m_i}$$

Now for objects which are relatively dense, air resistance is very small at low speeds and can be neglected compared to the weight. In such circumstances, the gravitational attraction alone is important, and we expect the acceleration of the object to be

$$\vec{\mathbf{a}} = \frac{m_g}{m_i}\,\vec{\mathbf{g}}$$

Here something remarkable appears. As we found out in Chapter 13, the ratio $\frac{m_g}{m_i}$ is the same for all objects. Consequently the acceleration does not depend on the mass of the object. In fact, by measuring m_i and m_g in terms of the same standard kilogram, we have made m_i equal to m_g; that is, $\frac{m_g}{m_i} = 1$. It then follows that for a freely falling body

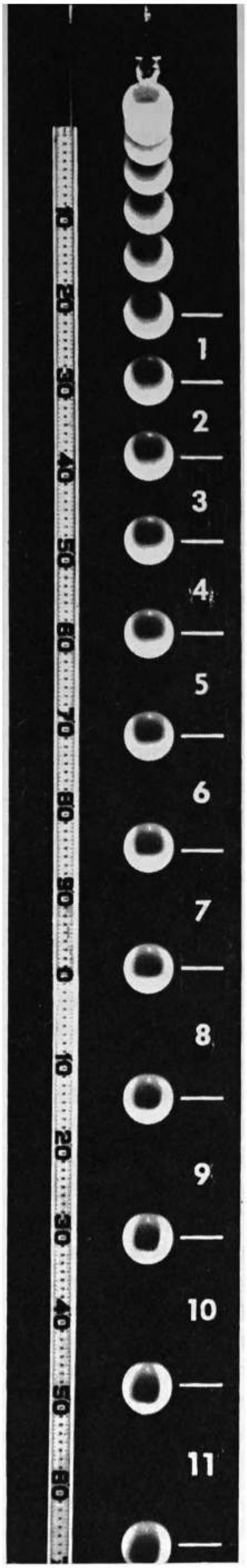
$$\vec{\mathbf{a}} = \vec{\mathbf{g}}$$

In other words, we predict that all objects moving under the influence of gravitation alone are accelerated in just the same way when located at the same place in the gravitational field. In addition, we should be able to measure the gravitational field strength g at any given location by observing the acceleration of any object as it moves through this position.

What do we find in fact? All bodies of sufficiently high density (large mass per unit volume) do fall with the same acceleration at the same place The magnitude of the acceleration with which they fall is very nearly 9.8 m/sec^2 anywhere near the surface of the earth. It makes no difference whether we measure it in a laboratory on the first floor of a building or on the top floor, whether the body starts from rest or has been projected with any reasonable vertical velocity.

Careful measurements of the multiple-exposure photograph of a falling object (Fig. 14 – 2) lead to the results shown in Table 2. The exposures are at $\frac{1}{30}$-sec intervals. These results illustrate the statement that the gravitational field near the earth's surface results in an acceleration of falling objects of about 9.8 m/sec^2 directed down.

We have limited our attention to compact, dense objects carefully selected to minimize the frictional resistance of the air. But if you drop a Ping-Pong ball it falls only a short distance



14–2 A flash photograph of a falling billiard ball. The position scale is in centimeters, and the time interval between successive positions of the ball is 1/30 second. This motion is analyzed in Table 2.

TABLE 2

INTERVAL NUMBER	DISPLACEMENT Δx(CM)	AVERAGE VELOCITY $\Delta x/\Delta t = v$(CM/SEC)	CHANGE IN AVERAGE VELOCITY Δv(CM/SEC)	ACCELERATION $\Delta v/\Delta t$(M/SEC^2)
1	7.70	231		
			32	9.6
2	8.75	263		
			31	9.3
3	9.80	294		
			32	9.6
4	10.85	326		
			34	10.2
5	11.99	360		
			33	9.9
6	13.09	393		
			32	9.6
7	14.18	425		
			32	9.6
8	15.22	457		
			32	9.6
9	16.31	489		
			35	10.5
10	17.45	524		
			32	9.6
11	18.52	556		
			Average acceleration	9.8

The analysis of the motion shown in Fig. 14 – 2. The calculated values of the acceleration are constant within the limits of accuracy of our measurements. Even though an accurate ruler was brought right up to the picture of the ball, the last significant figure in the Δx column is quite uncertain and is just a reasonable estimate of a fraction of a millimeter. It has been retained, however, to reduce successive errors that would accumulate if we rounded off early. Notice that we have kept only three significant figures in the velocity column.

before the force of air resistance balances the force of gravity, and the ball moves at constant velocity. In general, air resistance becomes greater with higher velocity. Therefore, if an object falls far enough, it will gain so much velocity that the air resistance becomes equal to the weight. Then the body continues to fall with constant velocity. This final constant velocity is called the terminal velocity of the falling body. (Problem: What happens if you throw a light bulb down faster than its terminal velocity?)

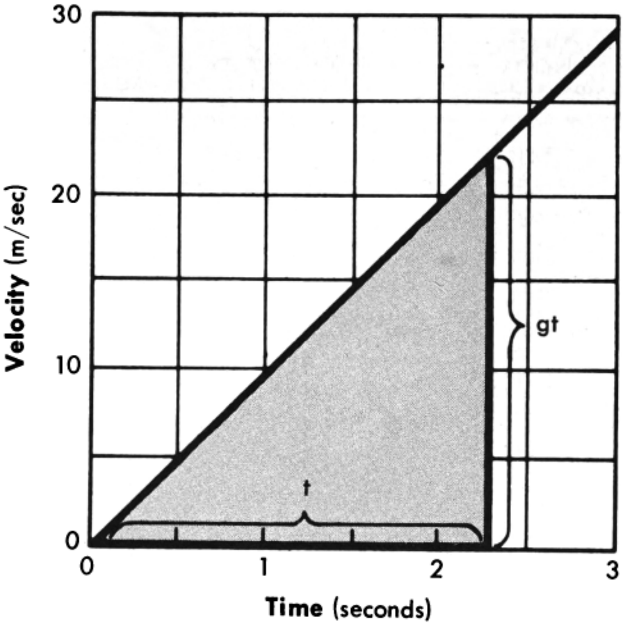
We can check that these deviations from free fall are indeed the result of air resistance by performing experiments in a vacuum. When we remove the air, we find that all objects, regardless of shape or density, fall with the same acceleration at a particular position near the earth's surface. Furthermore, because $\vec{g}$ does not change direction or magnitude appreciably unless we move through distances comparable to the size of the earth, the acceleration is closely the same for objects falling anywhere within a room, within a building, a city, or even a small country. In the region within which the gravitational field $\vec{g}$ is effectively constant and where gravitation alone is important, all objects fall with constant acceleration equal to $\vec{g}$. Starting from rest, in time t they pick up *downward* velocity

$$v = gt$$

and move down through the distance

$$d = \tfrac{1}{2}gt^2$$

given by the area under the graph of velocity versus time (Fig. 14 – 3). By using the flash photo of Fig. 14 – 2 again, you can show that the distance d moved from rest increases as $\frac{1}{2}gt^2$.



14–3 A velocity-time graph of a freely falling body. The vertical displacement, or distance fallen, is given by the area under the curve – the area of the triangle with base t and height gt. This area is $1/2gt^2$.

14–3 PROJECTILE MOTION: THE VECTOR NATURE OF NEWTON'S LAW OF MOTION

Objects that drop straight down under the influence of gravitational attraction alone all accelerate at the same rate. Do all objects also accelerate at that same rate when they move in other directions in the gravitational field? Figure 14 – 4 shows a series of flash photographs of two balls. The first ball started falling from rest at the moment that we projected the second one horizontally. We see that the vertical motions of the two balls are identical despite the fact that the horizontal motions differ. Also we see that the horizontal motion is at constant horizontal velocity, like motion when there is no force. The presence of the downward force does not change the horizontal motion; and the existence of horizontal motion does not change the effect of the downward force on the vertical motion. Our observation shows that the horizontal and vertical motions are independent. Each one follows from the appropriate component vector of the force: $\vec{F}_v = m\vec{g}$, and the vertical motion is the same as free fall; $\vec{F}_h = 0$, and the horizontal motion is without acceleration.

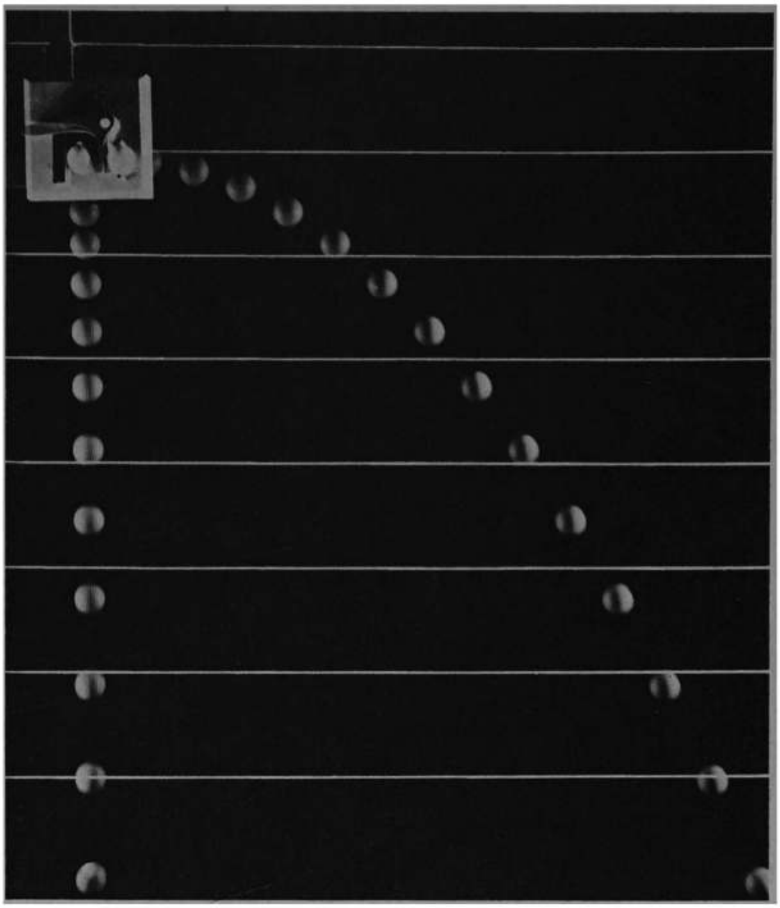
At the end of the last chapter we found that the net force $\vec{F}$ acting on a body is a vector. As we know from Chapter 12, the acceleration $\vec{a}$ is also a vector. This suggested that Newton's law of motion is a vector law. In addition, when the force $\vec{F}$ acts in the direction of motion, the acceleration $\vec{a}$ is related to the force by $\vec{F} = m\vec{a}$. In the last chapter, however, we did not investigate any motion in which the net force $\vec{F}$ acted at an angle to the velocity $\vec{v}$. We might, therefore, ask the question: Is Newton's law of motion valid when $\vec{F}$ and $\vec{v}$ are in different directions? Is the acceleration still in the direction of the force? Is the magnitude of $\vec{a}$ the same for the same size force?

Projectile motion is the first case we have examined in which $\vec{F}$ and $\vec{v}$ are in different directions. The observed fact is that the vertical gravitational force $\vec{F}$ produces the same vertical acceleration no matter whether there is horizontal motion or not. Although this observation is not enough to prove that $\vec{F} = m\vec{a}$, it provides supporting evidence for the idea that Newton's law in this simple form holds regardless of the direction of motion. After discussing the path of a projectile, we shall examine experiments with other forces acting across the direction of motion. We shall see that the vector law $\vec{F} = m\vec{a}$ also holds for those experiments.

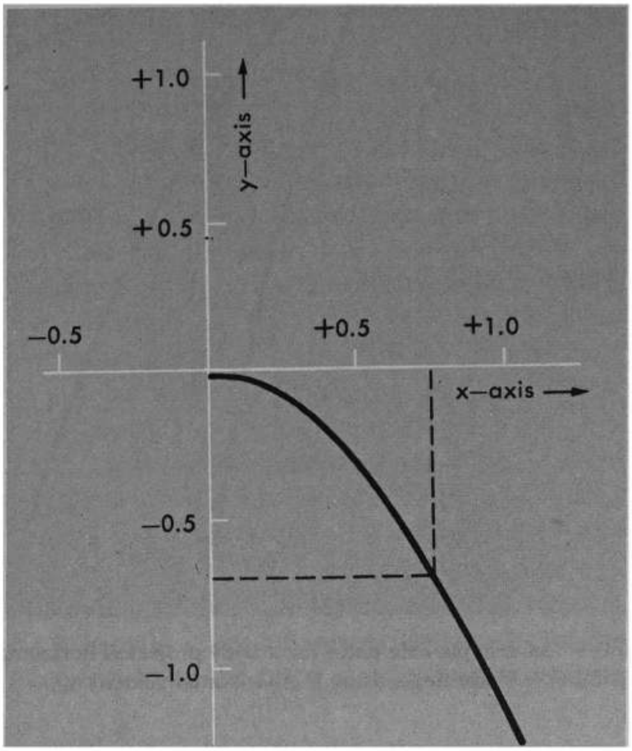
14–4 PROJECTILE MOTION: DETERMINATION OF THE PATH

In studying the motion of a projectile we encounter a new problem. Previously in our study of dynamics, we considered objects moving in a straight line under the action of a net force along the line of motion. The second ball in Fig. 14 – 4, however, traveled in a curved path. One of the important problems in dynamics is the determination of the path in such a situation.

If we know the position and velocity of an object at one moment, the path it follows as a function of time can be found from the force on the object and Newton's law of motion. For the projectile, however, we do not need to go all the way back to Newton's law and the gravitational force. We can find the path by combining the known vertical and horizontal motions. (As we saw in the last section, they go on independently in agreement with Newton's law and the known gravitational force.) For this purpose we choose a horizontal axis of reference (the x axis) and a vertical reference axis (the y axis) so placed that the origin ($x = 0$, $y = 0$) is at the point where the body is projected (Fig. 14 – 5).



14–4 A flash photograph of two golf balls, one projected horizontally at the same time that the other was dropped. The strings are 15 cm apart, and the interval between flashes was 1/30 second. This photograph is the same as Fig. 12–17.



14–5 The path of the second ball in Fig. 14–4 is plotted on a pair of coordinate axes. The scales on the figure measure the distance along the coordinate axes in meters. The x coordinate of the ball at any time t is v_0t (where $v_0 = 2$ m/sec), and its y coordinate is $-1/2gt^2$. For example when $t = 0.38$ sec, then $x = 0.75$ m and $y = -0.7$ m.

When the ball is projected with horizontal velocity v_0, we know that it continues to move along the x direction at this velocity. After a time t, the x coordinate of the position of the ball is therefore $x = v_0 t$. We also know that the vertical motion is the same as free fall. The y coordinate at the same time t is therefore $y = -\frac{1}{2}gt^2$. (The minus sign just tells us that the ball goes down rather than up.) These equations contain all the information about bodies projected horizontally with an initial velocity equal to v_0. The common value of the time t in the equations relates them to the motion of one single body rather than to the motion of two different bodies.

The path which the body follows is a curve, and we can express this curve by an equation relating the vertical position y to the horizontal position x at the same instant of time. To find this equation we eliminate the time t from the two equations $y = -\frac{1}{2}gt^2$ and $x = v_0 t$. From the second equation we see that $t = x/v_0$; and putting this expression for t into the first equation, we get

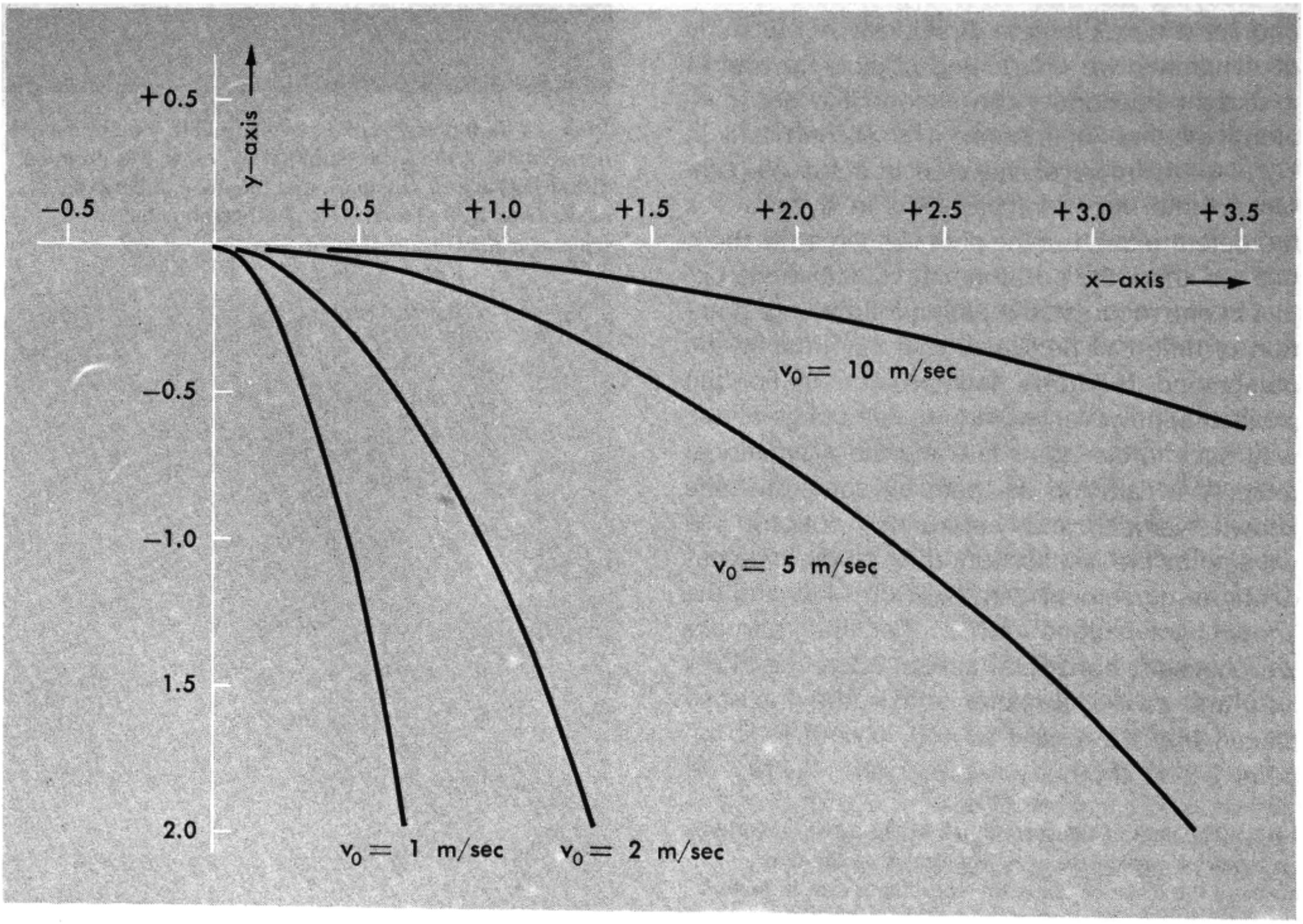
$$y = -\frac{1}{2}gt^2 = -\frac{1}{2}g\left(\frac{x}{v_0}\right)^2 = -\frac{g}{2v_0^2}\,x^2$$

The equation $$y = -\frac{gx^2}{2v_0^2}$$

is the equation of the path of the object. As shown in Fig. 14 – 5, the path is a parabola with its vertex at the place where the object is moving horizontally.

In Fig. 14 – 6 we have plotted several possible paths which correspond to different values of the initial horizontal velocity v_0. As we see, when the horizontal velocity is large, the parabola is rather flat. The projectile moves a long way sidewise before falling any great distance. On the other hand, small values of v_0 give sharply curved parabolas. The projectile moves a shorter distance horizontally in the time it takes to fall a given amount.

Here we have analyzed the problem of a



14–6 Several possible paths for a body projected horizontally. Note that the parabola's shape depends on the horizontal velocity v_0.

projectile which is fired horizontally. The more general case of a projectile fired with an initial velocity $\vec{v}_0$ at any angle with the horizontal can be handled in the same way. We again use the fact that the vertical and horizontal motions are independent. From the initial velocity vector $\vec{v}_0$, we find the initial horizontal and vertical components. The horizontal component of velocity never changes, and the vertical component undergoes a uniform change at the rate

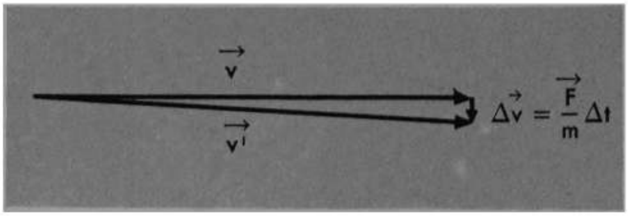
$$\frac{\Delta v_y}{\Delta t} = -g$$

When we work through the details the result is again a parabolic path with a flat region whose extent is determined by the horizontal velocity. The only difference is that the top of the parabola is not at the starting point of the motion.

What we learned in this section on projectile motion resembles what we learned earlier in Chapter 12. (Figure 14 – 4 is exactly the same as Fig. 12 – 17.) Yet there is an important difference. In earlier chapters we dealt only with the description of motion (kinematics) whereas here we considered the force which caused the motion (dynamics). The gravitational force $m\vec{g}$ and Newton's law in vector form allow us to predict as well as describe what happens. In our study of the dynamics of other motions we shall lean heavily on kinematic descriptions given previously. If you review Sections 12 – 6 to 12 – 8 at this time, the next sections of this chapter should be easier to understand.

14–5 DEFLECTING FORCES AND CIRCULAR MOTION

An object moves on a curved path because the force that pushes it has a component perpendicular to the direction of motion. The component of force along the path changes the speed, but it does not change the direction. On the other hand, a force perpendicular to the motion shoves the object sidewise so that the path curves. It leads to a perpendicular acceleration which changes the direction without changing the magnitude of the velocity vector as shown in Fig. 14 – 7.



14–7 The change in the velocity of a body when a force acts perpendicular to its path. Such a force changes the direction of the motion, but not the speed. The magnitude of $\vec{v}'$, the velocity after the force $\vec{F}$ has acted for the very short time Δt, is the same as that of $\vec{v}$.

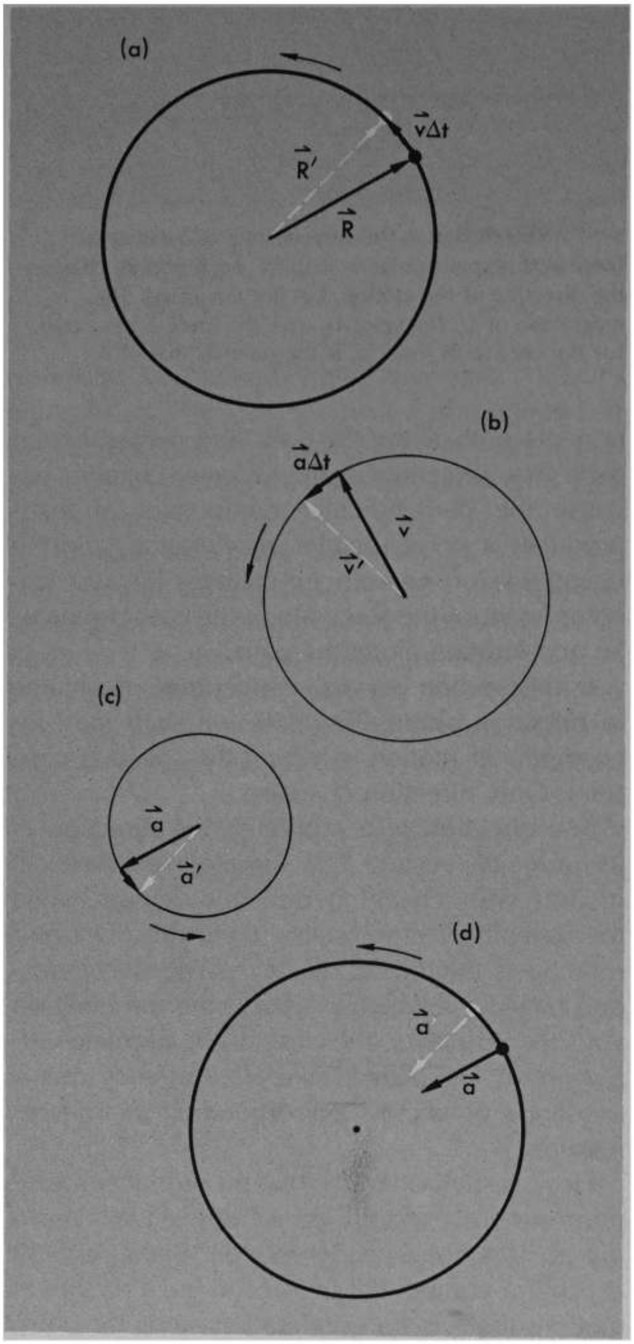
While a projectile is moving on its parabolic path, the force of gravity acting on it has components both along the path and perpendicular to it. The direction of the velocity changes because the perpendicular component of force produces a perpendicular acceleration, and the magnitude of the velocity changes because the component of the force along the path produces an acceleration along the path.

In this section we shall concentrate on change in direction alone. Therefore we shall look for examples of motion in which the speed is constant. Only direction changes.

The direction of a motion is the direction of its velocity vector. The simplest example of motion with changing direction occurs when the velocity vector rotates uniformly. Uniform rotation is familiar to all of us. The sweep second hand of a watch or clock, the turntable on your record player, the earth itself, all rotate uniformly. In Fig. 12 – 27 we have already met a sequence of vectors corresponding to uniform rotation.

How is such uniform rotation produced? Suppose we push steadily on a body, always pushing at right angles to its motion. Because there is never a component of force in the direction of motion, there is no acceleration along the path; the speed of the body stays constant while the direction of motion changes. Now if we keep the magnitude of the force constant, the direction of the path will change by the same amount in every equal time interval, and the path must be a circle. Each time the body moves at constant speed around the circle, the velocity vector rotates uniformly through 360 degrees.

Thus we have reached the important conclusion that a deflecting force of constant size perpendicular to the motion makes a body move in a circle with constant speed. The force and the acceleration of the body are directed toward the



14–8 **(a) The path of a body moving in a circle at constant speed. Its position vector $\vec{R}$ rotates uniformly. During a complete period T, the body moves around the circle a distance $vT = 2\pi R$. (b) The changing velocity vector of the body. $\vec{v}$ rotates at the same speed as $\vec{R}$, but at an angle of 90° to it. In a short time interval Δ_t, the velocity changes an amount $a\Delta t$. In one revolution, the tip of $\vec{v}$ moves once around the circle so $aT = 2\pi v$. Part (c) shows the changing acceleration vector, which is always perpendicular to the velocity. Part (d) again shows the circular path of the body. Here we have drawn the acceleration vectors in the same direction as in (c), but with their tails on the body. The acceleration vector always points from the body toward the center of the circle in which it moves.**

center of the circle, and the velocity vector is tangent to the circle at every point. The position vector drawn from the center of the circle is of constant length and equal to the radius of the circle (Fig. 14 – 8).

We shall now obtain two useful mathematical formulas for the acceleration in this circular motion. Since the motion occurs with constant speed v, and the body goes around the circumference $2\pi R$ of the circle in the time T,

$$v = \frac{2\pi R}{T}$$

The time T is called the period of the motion. In the same time T the direction of motion has also turned completely around and now points the way it did originally. The direction of motion rotates around steadily, always staying perpendicular to the radius, but the speed does not change. Consequently the velocity vector just turns steadily, with the same period T. (See Fig. 14 – 8.)

In Fig. 14 – 8 (b) we see (as we already noted in Chapter 12) that the velocity is carried around by an acceleration perpendicular to it—just as the radius vector is carried around by a velocity perpendicular to it. The tip of the velocity vector therefore goes around a circle of radius v and circumference $2\pi v$ in the time T; and the magnitude of the acceleration which produces this change is

$$a = \frac{2\pi v}{T}$$

By combining the last two equations, we can eliminate either T or v, expressing the acceleration alternatively as

$$a = \frac{v^2}{R}$$

or as

$$a = \frac{4\pi^2 R}{T^2}$$

These expressions will be valuable in studying satellite motion, the planetary system, and atomic physics.

The acceleration vector is perpendicular to the velocity vector; and the velocity vector is perpendicular to the position vector $\vec{\mathbf{R}}$. The two right angles add up to 180°. Therefore, the direction of the acceleration vector is opposite to that of $\vec{\mathbf{R}}$. (See Fig. 14 – 8.) Using this information about the direction of $\vec{\mathbf{a}}$ and the last formula for its magnitude, we can write the vector relation

$$\vec{\mathbf{a}} = -\frac{4\pi^2\vec{\mathbf{R}}}{T^2}$$

If we imagine $\vec{\mathbf{a}}$ with its tail on the body, we shall see that it points from the body toward the center of the curve around which the body moves. For this reason the acceleration is called *centripetal* (center-seeking) *acceleration.*

Now let's go back to the force that produces this acceleration. Newton's law $\vec{\mathbf{F}} = m\vec{\mathbf{a}}$ tells us how the force and acceleration are related. Therefore, with our mathematical expression for the acceleration in terms of the radius R of the motion and its period T, we can relate the force to the circular motion it produces. The result is

$$\vec{\mathbf{F}} = m\vec{\mathbf{a}} = -\frac{m4\pi^2\vec{\mathbf{R}}}{T^2}$$

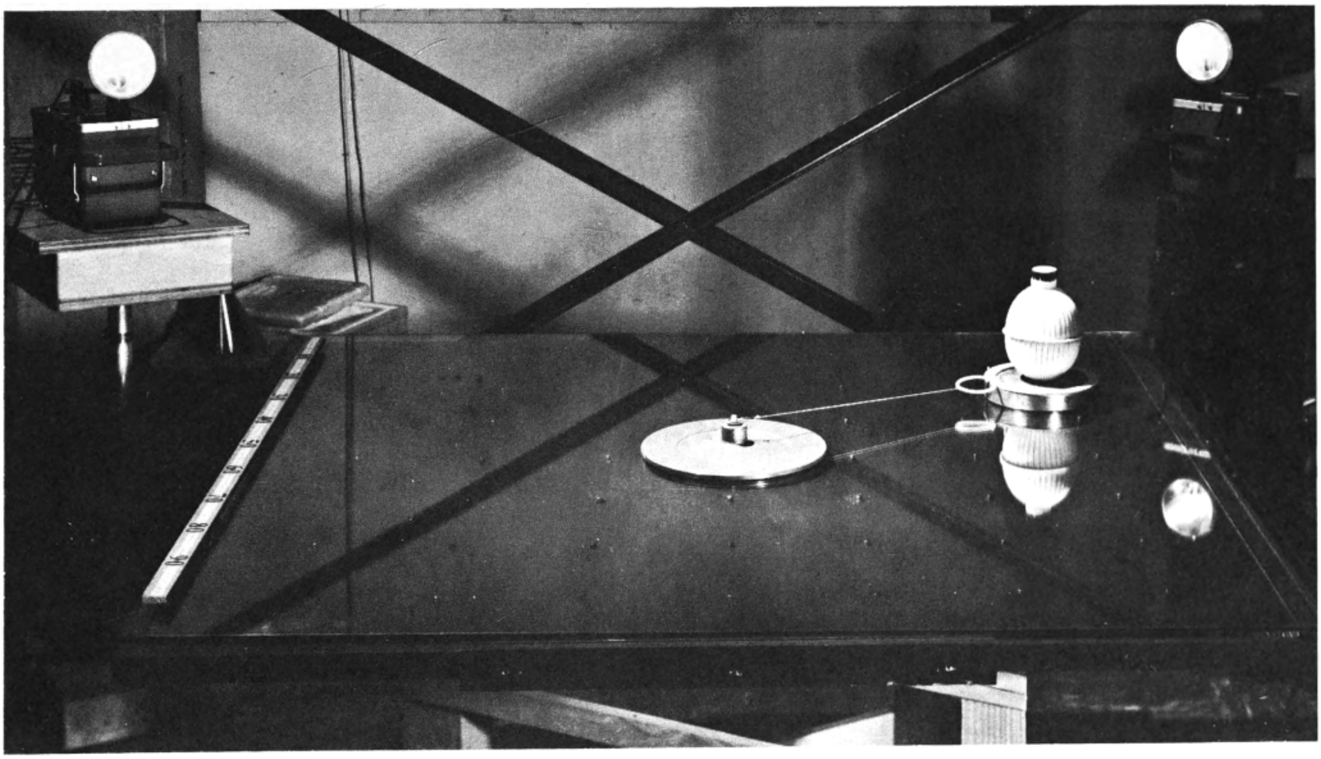
The magnitude of the force can also be related to the speed v and radius R as

$$F = \frac{mv^2}{R}$$

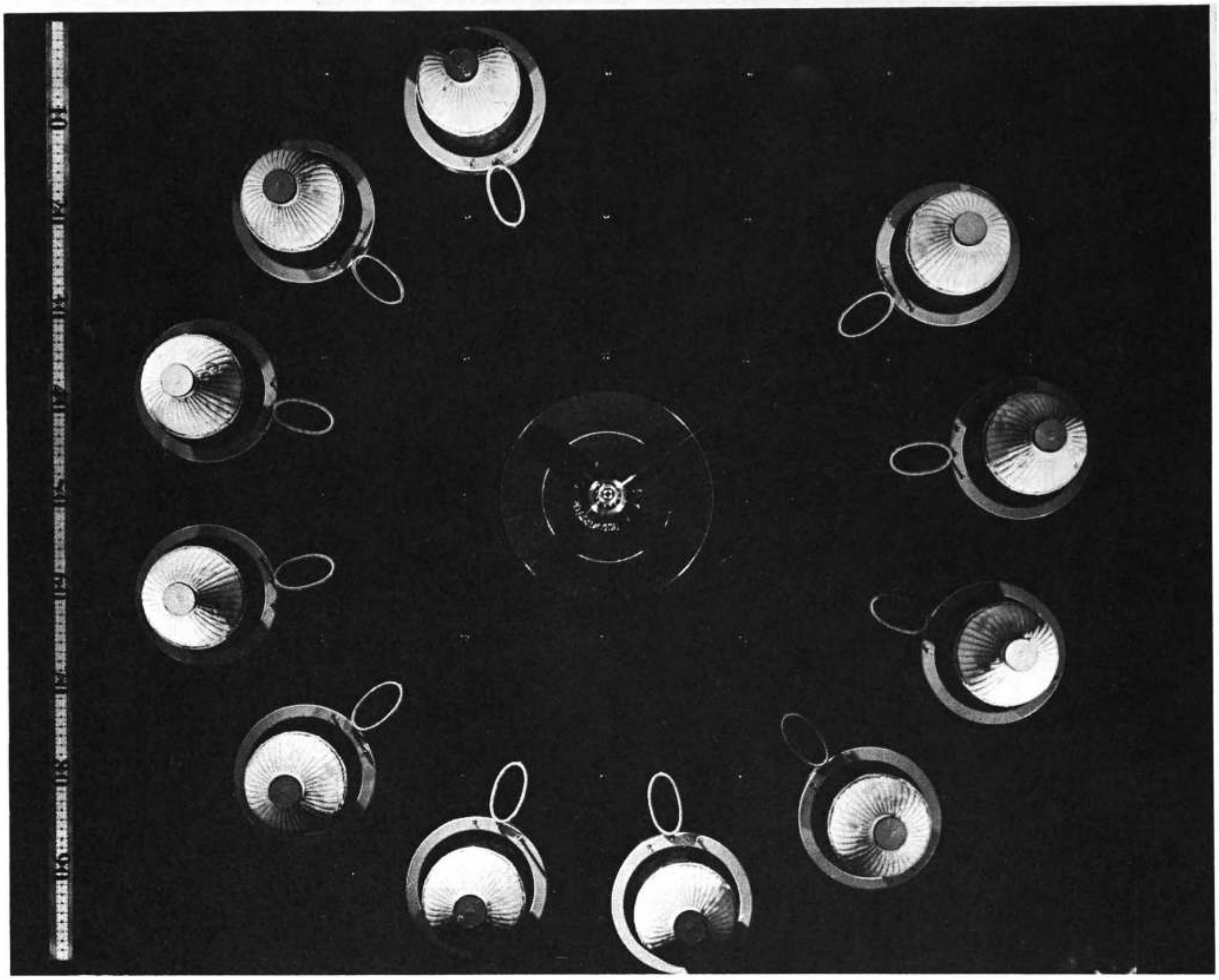
Remember that the direction of $\vec{\mathbf{F}}$ is inward, toward the center of the circle, and the force is called *centripetal* force. The last equation actually holds for any path where F is the perpendicular deflecting force and R the radius of curvature (the radius of a circle which just fits the actual path near the point in question).

We can check our equations experimentally. Figure 14 – 9 shows an apparatus designed to do this with a Dry Ice puck. The puck rests on a horizontal glass table. One end of the string is attached to a bearing at the center of the table, and the other end is connected to a spring loop fastened to the puck.

In the experiment the puck was given just the right push to make it move around a circle. Figure 14 – 10 is a stroboscopic flash photograph of the motion. The light flashed 2.4 times a second and was stopped before one complete revolution was registered. The spacing between successive pictures of the puck shows that the speed was essentially constant. The loop on the puck is stretched, as you see by comparing it with Fig. 14 – 9. The stretch stays the same



14–9 An arrangement for measuring the centripetal force acting on a body moving in a circular path. The puck is at rest; no force is acting; the loop on the puck is not extended.



14–10 A puck moving in a circle at uniform speed. Note that the loop is extended the same amount at each position, indicating that a constant force is acting.

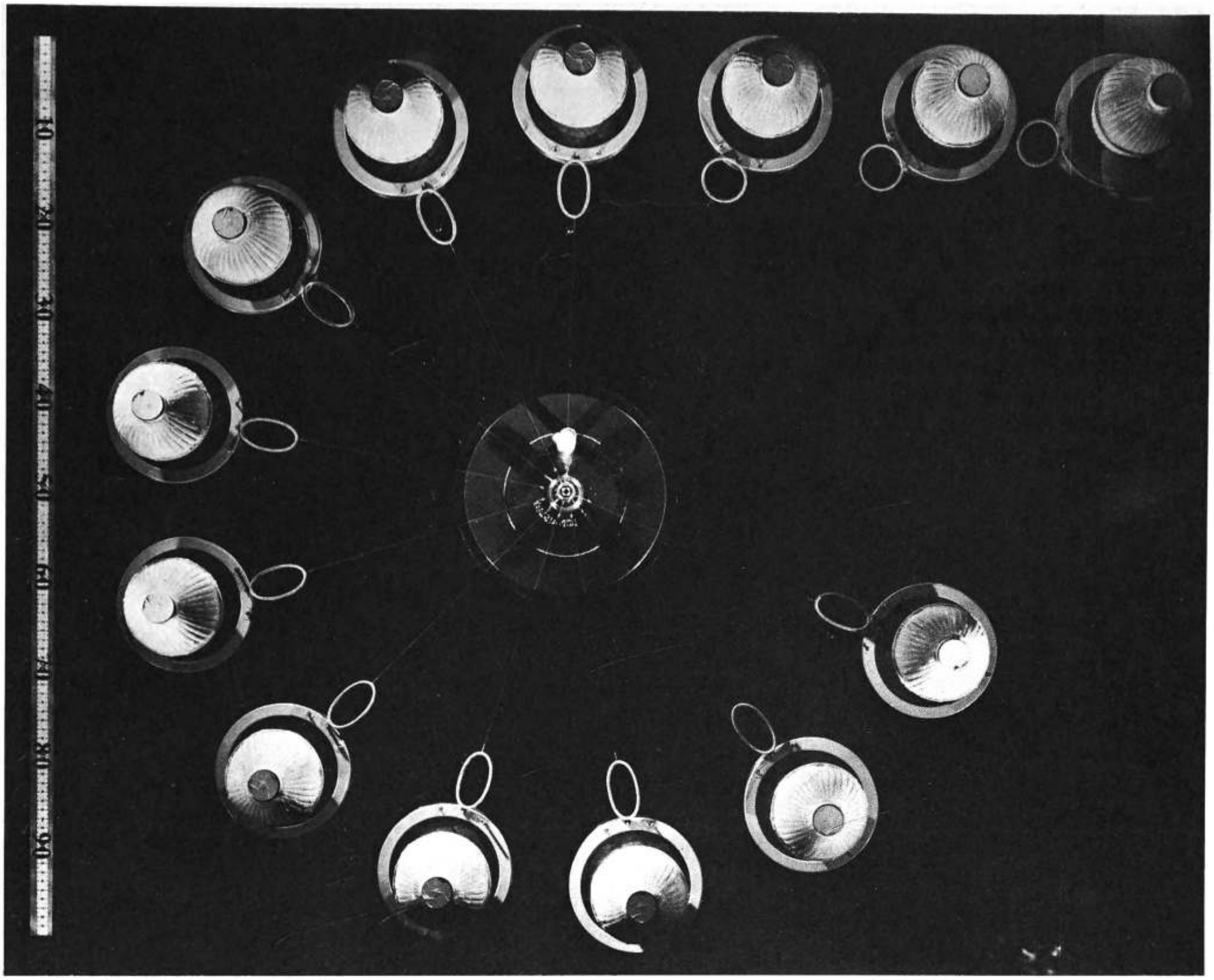
throughout the motion, and from it we can find the force which makes the puck run around in a circle.

The measured extension of the loop tells us the force is 2.4 newtons. The time taken to move through 10 intervals from the first position to the last is 10 intervals $\times \frac{1}{2.4}$ sec/interval. In this time the puck moved through a measured angle of 286°. Therefore the time T for one complete revolution is $10 \times \frac{1}{2.4} \times \frac{360°}{286°} = 5.2$ sec. The radius was 0.44 meter, and the mass of the puck and loop was 3.9 kg. Using these values in our expression for the magnitude of the centripetal force F which should make the puck go around the circle, we get

$$F = \frac{m4\pi^2R}{T^2} = \frac{(3.9\text{kg})4\pi^2(.44\text{m})}{(5.2\text{sec})^2} = 2.5 \text{ nt}$$

We see that this experiment confirms our expression for the centripetal force to within the accuracy of our measurements.

Suppose the string breaks. The puck no longer moves in a circle, since the centripetal force has been removed. Since no external force now acts on the puck, *it moves in a straight line at constant speed.* This motion is shown in Fig. 14 – 11, where the string was burned through by a gas torch. Notice that after the string breaks, the puck moves along a tangent to the original circle; it does not move out along a radius.



14–11 **A puck in circular motion photographed at a flash rate of 2.4 per second. When the puck reached the top of the photograph, the string attached to it was burned with a torch. The puck continued to move in a straight line with the velocity it had when the string broke.**

14–6 EARTH SATELLITES

Only for distances small compared with the earth's radius is the weight of a body constant in direction and magnitude. When we throw a projectile as fast as is now possible with rockets, it may travel so far that we can no longer neglect the change in direction of the force of gravity. Because the earth is almost spherical, we conclude that the force of gravity always points toward its center. If we look from outer space, we see that the direction of the force is completely different for different points around the earth (Fig. 14 – 1).

If a projectile is launched horizontally with the right initial velocity, it will move at constant speed in a circular orbit around the earth. However, since it is very difficult to project an object in exactly the right direction, and with just the right speed, artificial satellites actually follow elliptical orbits (see next chapter). We shall analyze only the special case of circular motion of a satellite, since it is relatively simple mathematically.

The earth satellites which have been successfully launched are multistage rockets. The last stage is fired almost horizontally at high altitude where air resistance is small. There is no essential difference in the dynamical behavior of a man-made satellite and one found in nature, such as the moon. But the radii of their orbits are different and, as a result, their periods of revolution (that is, the times for one trip around the earth) are different.

Let us compute the period of revolution of a satellite moving around the earth. For motion in a circle the magnitude of the acceleration is $a = v^2/R$, and since the centripetal force is the gravitational attraction of the earth, the magnitude of this acceleration must be equal to g, the gravitational field strength. Hence the speed of the satellite is given by

$$\frac{v^2}{R} = g \quad \text{or} \quad v^2 = gR$$

where R is the radius of the circle and g is the value of the acceleration of gravity where the satellite is. Suppose the satellite is 400 km above the earth. Then R = radius of earth + 400 km = 6.8×10^6 meters; and g is about 8.6 m/sec^2 out there. Therefore

$$v^2 = gR = 8.6 \times 6.8 \times 10^6 \text{ m}^2/\text{sec}^2$$

and $$v = 7.6 \times 10^3 \text{ m/sec}$$

This is a speed of about 5 mi/sec or 18,000 miles per hour.

To get the period T, we note that the circumference of the circle ($2\pi R$) is the distance covered in one revolution at constant speed v. Therefore

$$2\pi R = vT \quad \text{or} \quad T = \frac{2\pi R}{v}$$

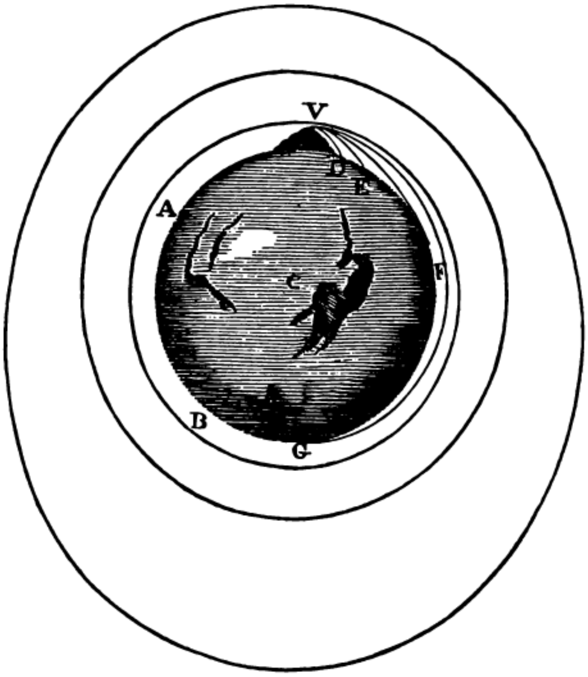
We can estimate the period of the satellite by substituting $R = 6.8 \times 10^6$ m and the value $v = 7.6 \times 10^3$ m/sec that we have just computed. The result is

$$T = 2\pi \frac{6.8 \times 10^6}{7.6 \times 10^3} = 5.6 \times 10^3 \text{ sec} = 93 \text{ min}$$

The calculations we have made for the period and velocity of an artificial satellite were needed to put the first artificial satellite in orbit. Newton suggested the possibility of launching a satellite by firing a projectile from a high mountain with a velocity of about 5 miles per second, and drew the diagram of Fig. 14 – 12 to explain his proposal. It is one of the triumphs of the development of the dynamics of Galileo, Newton, and others that it was possible to discuss such projects with confidence in their ultimate possibility.

Why, then, was the first artificial satellite not placed in orbit in the seventeenth century? You can guess the answer. There were no guns or rockets powerful enough. Man's understanding of broad scientific laws often runs ahead of technology. The detailed application of scientific knowledge requires much time and labor.

Sometimes it is the other way round—technology runs ahead of science. Now that our technology has enabled us to establish artificial satellites, we shall gain new observations of the universe. Some of these observations have previously been denied us by the curtain of the earth's atmosphere. The new data will give us new knowledge of cosmic rays and of the density of matter in interplanetary space, and will help shape our ideas of the universe. In the long run, basic knowledge and technological applications go hand in hand—one helps the other.



14–12 A drawing from Newton's "System of the World" (attached to later editions of the "Principia"), showing the paths that a body would follow if projected with various speeds from a high mountain. As you can see, Newton was aware that a body would go into an orbit around the earth if its speed was great enough. The orbits from V ending at D, E, F, and G are for projectiles thrown with higher and higher horizontal velocities. Newton recognized that air resistance would prevent the motion of satellites close to the earth from following his ideal paths or from continuing for a long time. He pointed out that satellites could move permanently in the outer orbits.

14–7 THE MOON'S MOTION

The moon is an earth satellite. We can compute its centripetal acceleration from the following observations. The period of the moon's motion is 27.3 days, or 2.3×10^6 sec; and the distance from the earth to the moon is about 3.8×10^8 meters (about 240,000 miles). The magnitude of the moon's acceleration toward the earth is

$$a = \frac{4\pi^2 R}{T^2} = \frac{4 \times \pi^2 \times 3.8 \times 10^8 \text{ m}}{(2.3 \times 10^6 \text{ sec})^2}$$
$$= 2.7 \times 10^{-3} \text{ m/sec}^2$$

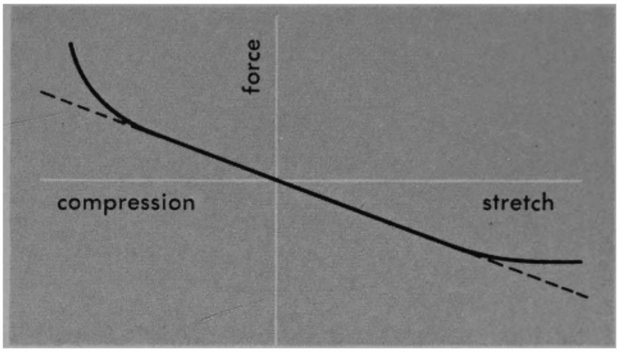
This acceleration is much smaller than the acceleration of a satellite near the earth's surface. Comparing it with $g = 9.8$ m/sec², we see that the gravitational attraction has fallen off by a factor of about 2.7×10^{-4}. This evidence of the weakening of gravitational attraction as the separation increases was one of the things that led Newton to his law of gravitation, as we shall see in the next chapter.

14–8 SIMPLE HARMONIC MOTION

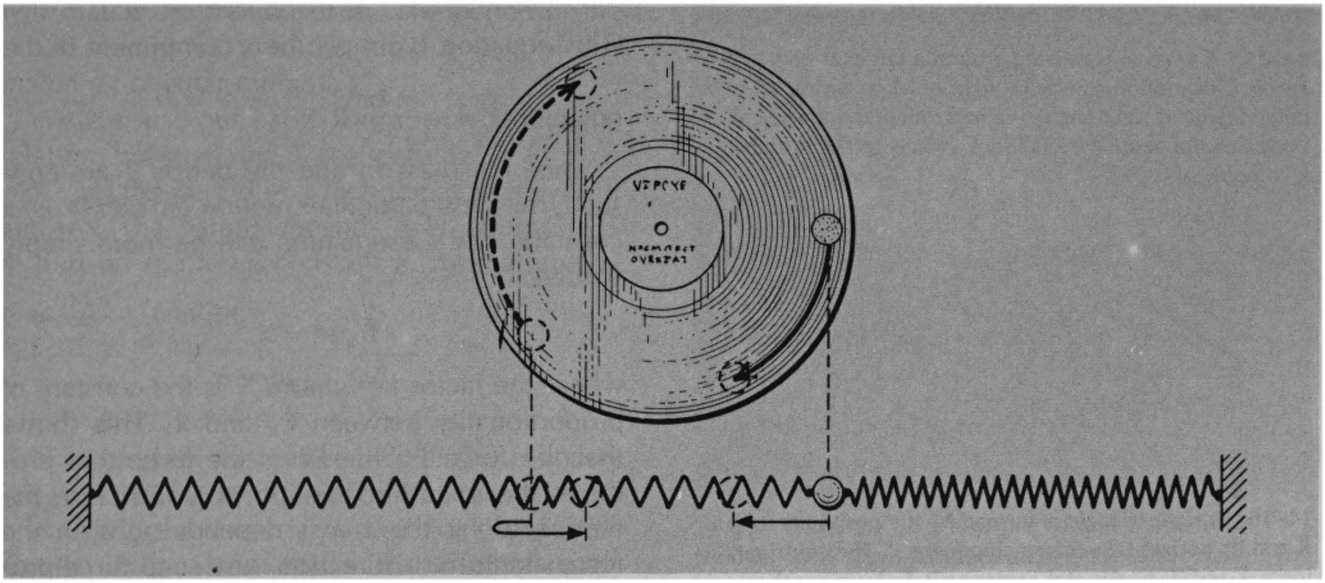
When we stretch a spring, it pulls back with a force proportional to the stretch. Also, when we compress a spring, it pushes against us with a force proportional to the compression (Fig. 14 – 13). If we attach a mass to a stretched spring and let go, it will oscillate to and fro owing to the force of the spring. For small stretch or compression (as long as we remain in the linear part of the graph), this force can be described by

$$\vec{\mathbf{F}} = -k\vec{\mathbf{x}}$$

Here k is a positive constant, and $\vec{\mathbf{x}}$ is the displacement of the mass from the position of no distortion of the spring (chosen positive for stretch, negative for compression). The minus sign in the equation indicates that $\vec{\mathbf{F}}$ is in the opposite direction from $\vec{\mathbf{x}}$; thus it is a restoring force which pulls the system back toward its equilibrium position.

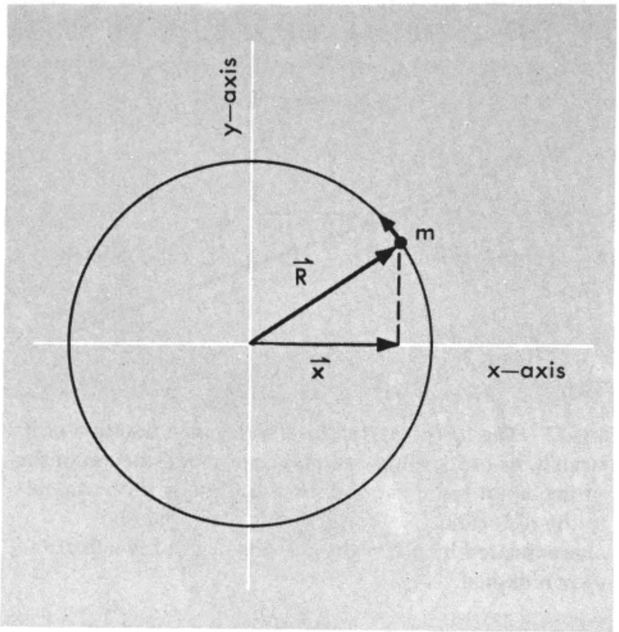


14–13 The force exerted by a spring as a function of its stretch, or of its compression. When the distortion of the spring is not too great, the force is directly proportional to the distortion. The curve in this region can be approximated by a straight line, the ends of which are shown dashed.

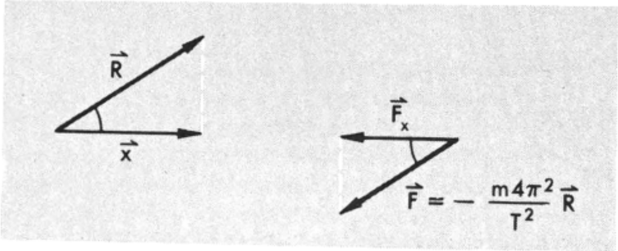


14–14 The to-and-fro motion of a mass on a spring can be matched by one component of the motion of a mass on a circular turntable. We must choose an appropriate radius for the turntable and turn it at the right speed.

Linear restoring forces such as these always lead to similar to-and-fro motions, called *simple harmonic motion.* Let us examine this motion in some detail. If we were to use Newton's law directly to predict the motion, we would be confronted with a mathematically complicated problem; but the motion can be derived from circular motion, as we shall now show. Circular motion in a plane looks just like the simple harmonic to-and-fro motion if we examine it from the edge of the plane. To get an idea of the motion, move your thumb steadily around a horizontal circle level with your eye. The straight-line motion you see is simple harmonic motion, as we can tell by matching such a motion with the motion of a mass on a spring (Fig. 14 – 14).



14–15 A mass m is moving around a circle at constant speed. This motion can be thought of as consisting of two components at right angles – a horizontal and a vertical vector component. We shall look closely at the horizontal x component.



14–16 A right triangle is formed by the position vector $\vec{R}$ and its horizontal vector component $\vec{x}$. The centripetal force $\vec{F}$ and its horizontal component $\vec{F}_x$ form another right triangle. The triangles are similar because the angle between $\vec{R}$ and $\vec{x}$ is equal to the angle between $\vec{F}$ and $\vec{F}_x$ (since $\vec{R}$ is parallel to $\vec{F}$, and $\vec{x}$ is parallel to $\vec{F}_x$).

To study this motion, we once again return to the fundamental property of vectors which allows us to represent any motion in a plane as the combination of two independent motions in fixed directions at right angles to each other. We shall describe circular motion at constant speed in terms of these rectangular components. We shall then examine one of these components in detail and also find the force associated with the motion described by that component.

Figure 14 – 15 shows a circle around which a body of mass *m* is moving with constant speed; let us examine the horizontal component of this motion. In Fig. 14 – 16 we have drawn a right triangle with the position vector $\vec{\mathbf{R}}$ and its vector component $\vec{\mathbf{x}}$ in the horizontal direction. We have also drawn the similar right triangle with the centripetal force $\vec{\mathbf{F}}$ and its vector component $\vec{\mathbf{F}}_x$. From these similar triangles we see that the magnitudes of $\vec{\mathbf{F}}_x$ and $\vec{\mathbf{x}}$ are in the same ratio as those of $\vec{\mathbf{F}}$ and $\vec{\mathbf{R}}$ and that $\vec{\mathbf{F}}_x$ and $\vec{\mathbf{x}}$ point in opposite directions.

Since $\vec{\mathbf{F}} = -\frac{m4\pi^2}{T^2}\vec{\mathbf{R}}$ (see Section 14 – 5), the relation between $\vec{\mathbf{x}}$ and $\vec{\mathbf{F}}_x$ is therefore

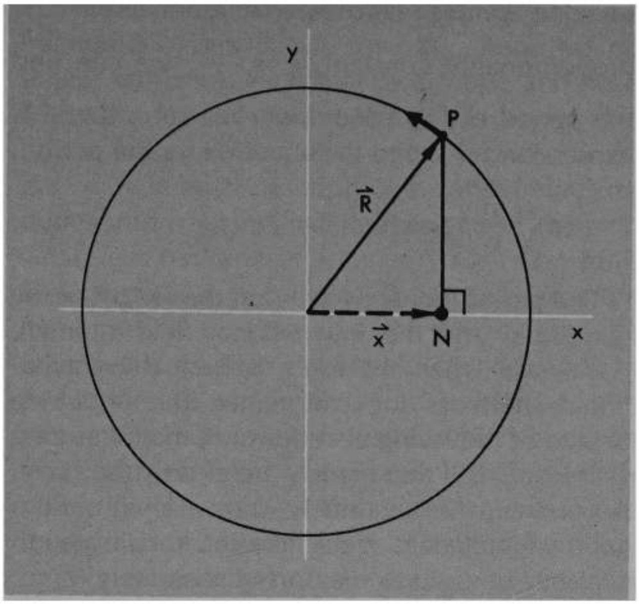
$$\vec{\mathbf{F}}_x = -\frac{m4\pi^2}{T^2}\vec{\mathbf{x}}$$

(This equation is simply the *x* component of the equation $\vec{\mathbf{F}} = -\frac{m4\pi^2}{T^2}\vec{\mathbf{R}}$.)

Since the mass *m* and the period *T* are constant for any particular motion, $m4\pi^2/T^2$ is a constant, and the equation can be more simply written

$$\vec{\mathbf{F}}_x = -k\vec{\mathbf{x}}$$

where the factor $k = m4\pi^2/T^2$ is the constant of proportionality between $\vec{\mathbf{F}}_x$ and $\vec{\mathbf{x}}$. This shows that the force $\vec{\mathbf{F}}_x = -k\vec{\mathbf{x}}$ is the force that produces the to-and-fro motion of $\vec{\mathbf{x}}$. Because the motion along the *x* axis depends only on the force along the *x* direction, any such force produces a motion exactly like the motion of point *N* at the end of $\vec{\mathbf{x}}$ in Fig. 14 – 17 when the point *P* in the figure goes steadily around the circle.



14–17 The mass at N was originally displaced from the origin by a distance R along the x axis. If it is acted on by a force such that $\vec{\mathbf{F}} = -k\vec{\mathbf{x}}$, then it will move back and forth along the x axis; its motion will be the same as the x component of the motion of a point P moving uniformly in a circle of radius R.

We have now linked up the *x* component of circular motion with the motion of any mass *m* when acted on by a restoring force $\vec{\mathbf{F}} = -k\vec{\mathbf{x}}$ no matter what provides the force, and this link enables us to calculate the period. If we have a mass *m* moving under a force $\vec{\mathbf{F}} = -k\vec{\mathbf{x}}$, we can always imagine a suitable circular motion that will match the motion of *m*. The period of the actual motion of *m* is the same as that of this matching circular motion.

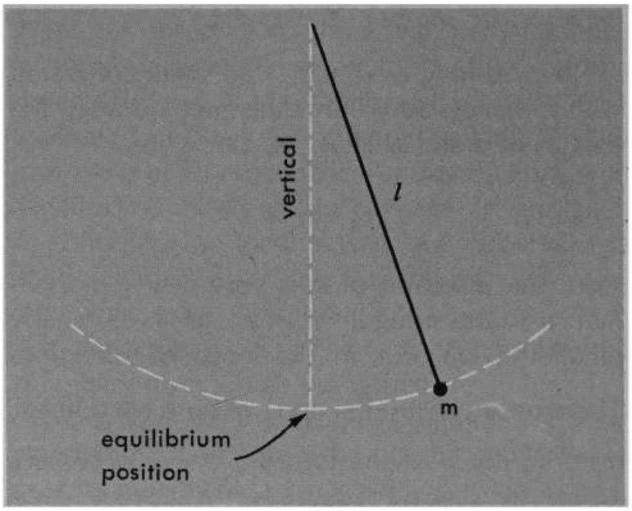
We can use our earlier discussion to find the relation between the force constant *k* in $\vec{\mathbf{F}} = -k\vec{\mathbf{x}}$ and the period in which the mass moves to and fro. The period depends only on the mass *m* and on the force constant *k*. By rearranging $k = \frac{m4\pi^2}{T^2}$, we find

$$T = 2\pi\sqrt{\frac{m}{k}}$$

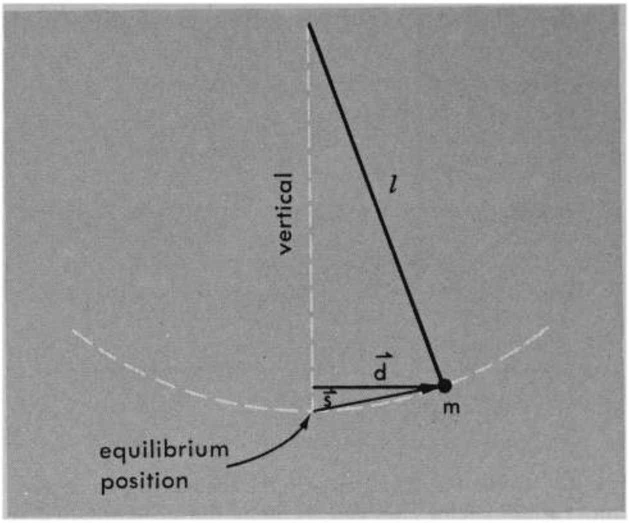
This expression is reasonable. If the restoring force increases rapidly with distance (that is, if *k* is large), the mass is pushed back and forth rapidly: *T* becomes small. On the other hand, if the mass is large it responds more sluggishly to the force; and so the period is greater the larger the mass. One result of our reasoning may seem astonishing: the period does *not* depend on the amplitude *R* of the motion. This is a result you can easily check experimentally. Try it.

A simple pendulum is just a body of mass *m* on the end of a string of length *l*. (Fig. 14 – 18.) We shall describe the displacement of the mass from its equilibrium position by the vector $\vec{\mathbf{s}}$. The vector component of $\vec{\mathbf{s}}$ in the horizontal direction is $\vec{\mathbf{d}}$ (Fig. 14 – 19); this measures the horizontal displacement of *m* from equilibrium.

The motion of the mass *m* is very nearly simple harmonic motion. To show this, we



14–18 A simple pendulum: a mass m on a string of length l moves along the arc of a circle.



14–19 When the mass is displaced only a little way from its equilibrium position, there is very little difference between the actual displacement $\vec{\mathbf{s}}$ and the horizontal component of the displacement $\vec{\mathbf{d}}$. Imagine that the displacement is much smaller than it is shown here, compared with the length of the string.

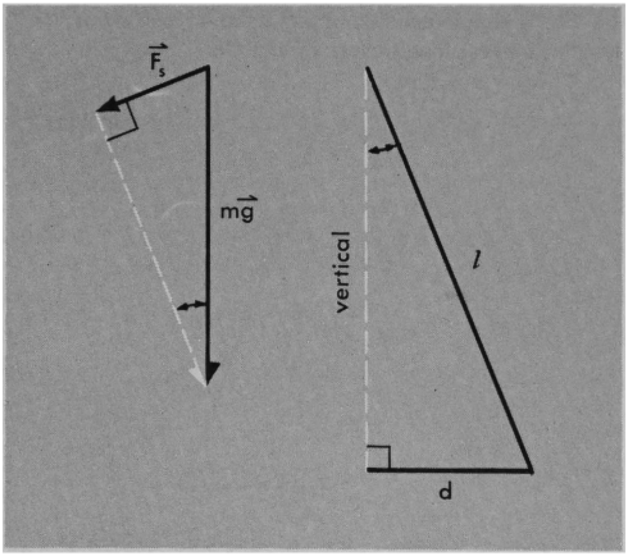
must show that there is a linear restoring force, a force which is proportional to the negative of the displacement. As we see in Fig. 14 – 20, the weight of the body $m\vec{g}$ can be broken into rectangular vector components. One component, shown dotted in the figure, is taken along the string and simply keeps the string pulled tight; the other component is perpendicular to the string, points along the circular path, and shoves the mass back toward the equilibrium position. From the similar triangles of Fig. 14 – 20, the magnitude F_s of this last component is given as

$$\frac{F_s}{mg} = \frac{d}{l} \quad \text{or} \quad F_s = \frac{mg}{l}d$$

Now, as long as d and s are small compared with l, there is very little difference between the lengths of d and s. Thus, we can write

$$F_s = \frac{mg}{l}s$$

Also, the direction of $\vec{\mathbf{s}}$ is very nearly perpendicular to the string if s is small, or is very nearly parallel to $\vec{\mathbf{F}}_s$ but in the opposite direction. Therefore $\vec{\mathbf{F}}_s = -\frac{mg}{l}\vec{\mathbf{s}}$ and we do have a linear restoring force of the form $\vec{\mathbf{F}}_s = -k\vec{\mathbf{s}}$ necessary to give simple harmonic motion. From the proportionality constant, $k = \frac{mg}{l}$, we can find the period of the pendulum. We substitute this expression for k into the equation for the period, $T = 2\pi\sqrt{\frac{m}{k}}$, and find $T = 2\pi\sqrt{\frac{l}{g}}$.



14–20 **In the left-hand triangle the weight $m\vec{g}$ has been broken into vector components along the string (dashed line) and perpendicular to the string ($\vec{F}_s$). The other triangle shows the string of length l, the horizontal displacement d of the mass from the vertical, and the vertical center line. The two right triangles are similar, since F_s is perpendicular to l and $m\vec{g}$ is perpendicular to d.**

The period depends only on the length of the pendulum and the gravitational field strength. As always when the force is basically gravitational, the mass does not matter. The amplitude or size of the swing also does not matter as long as it is so small that d and s are closely the same. A pendulum of accurately known length and a good independent clock like the rotating earth can thus be used to measure g accurately.

14–9 EXPERIMENTAL FRAMES OF REFERENCE

Turn around. Apparently the room rotates the other way. The corner of the room goes around a circle; so does a marble on the table. But no force mv^2/r on the marble is required to accelerate the marble while it moves in its apparently circular path. Under the action of no net force the marble is simply at rest when we describe it with respect to the room, but it appears to be accelerating when you describe its motion with respect to yourself.

Suppose you are in a car that is accelerating in a straight line. You look out and see a ball flying through the air. It appears to curve violently in the horizontal plane, although you realize that it moves on a straight path over the ground. If you describe the motion of the ball with respect to the car, the ball undergoes an acceleration which does not correspond to any sensible forces.

In Chapter 12 frames of reference were discussed in connection with the description of motion—that is, in connection with kinematics. The essential point made there is that any frame of reference can be used to describe a motion truly. So the sensible thing to do is to use the frame of reference that provides the *simplest* description. In other words, we have complete freedom of choice and use this freedom to make life as easy as possible.

But how about dynamics? Have we the same freedom? It is sad but true that here we no longer can use every reference frame and keep Newton's law of motion. Newton's law of motion is *not* correct in all frames of reference. To give a dynamical description in some frames requires more complicated laws. How then do we select reference frames in which Newton's law of motion is correct? The answer is: from experimental observations. Remember how we were led to Galileo's principle of inertia and to the statement of Newton's law of motion; and remember that any conclusion which is drawn from experimental observation is true only within the accuracy of the measurements. Since we arrived at Newton's law of motion from laboratory experiments on the earth, we know that a frame of reference rigidly attached to the earth is a good one—at least within the accuracy of our measurements. On the other hand, an accelerating car is not a frame of reference in which Newton's laws hold.

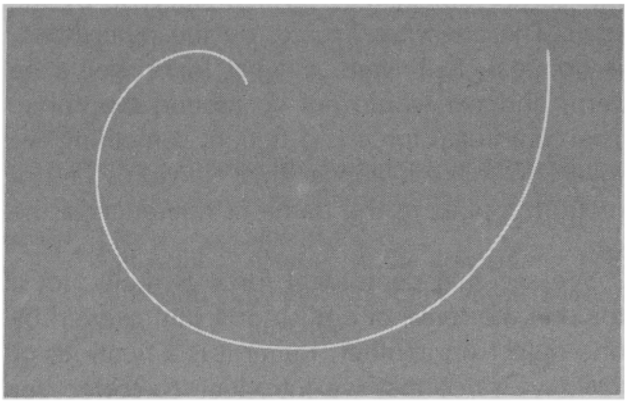
14–10 FICTITIOUS FORCES IN ACCELERATED FRAMES

We are all aware of the strange forces which appear when motion is observed in some frames of reference. If we ride in a closed car at constant speed on a straight, smooth road, no forces act on us except those which act when we sit in a chair at rest. But when you drive around a curve, especially at high speed, you are conscious of a force: the door of the car pushes on your body. If you think of this experience in the frame of reference of the car, it appears to be a cause for concern. A force acts on you, but you do not move. You are likely to say that another force pushes you against the door and holds you there. This new centrifugal force acting away from the center of the curve just balances the force exerted by the door. In this way you may explain why you don't move with respect to the car. But here is a real dilemma: how can a change in the motion of a reference system (the car) actually create new forces? Is this centrifugal force real?

To answer these questions we start by looking at a simple experiment. We take a Dry Ice puck, for example, hold it at rest on a flat, frictionless surface, and let it go. What happens? Exactly nothing. The puck stays at rest, and we conclude from this observation that no net force acts on it. In particular, no horizontal force acts on the ice puck.

Now we perform the same experiment on a frictionless table located on a large, rotating merry-go-round. Our whole laboratory rotates uniformly with respect to the earth. Now what do we observe when we let the puck go? The puck does *not* stay at rest but moves *relative to us*—that is, relative to the merry-go-round—in a curve of the sort seen in Fig. 14 – 21. It is called an involute. Now we are really in trouble! Here our observations show that, in the absence of forces, a body moves in an involute, not in a straight line with constant speed. What a contradiction of Galileo's principle of inertia! How can we properly reconcile our observations with Galileo's law? We might say: we are mistaken to think that no force acts on the puck; such a motion shows that there *is* a complicated force. But what exerts this force? Nothing of which we are aware. So we are right back in the quandary from which we started.

Now look at the whole situation from the standpoint of an observer on the ground. He sees us moving in a circle with constant speed. We are holding an ice puck which moves in the circle with us. When we let go of the puck, it moves in a straight line, tangent to the circle, at constant speed, as Galileo's principle says it must. The man on the ground understands our dilemma and points out to us that our troubles have come, not from a breakdown of the laws of dynamics, but from the fact that we made our



14–21 If we release a puck on a frictionless table when we are in a room on a large merry-go-round; it seems to move on a curved path. Such a path is called an involute.

observations in a rotating system of reference. This rotating system is an accelerated system, and we can solve all our problems by allowing for the motion of our rotating reference system. Now we know the cause of our trouble. The apparent violation of Galileo's principle and Newton's law of motion is due to the *acceleration* of the rotating frame of reference relative to the ground.

Thus if the observer on the merry-go-round holds the puck at rest relative to him, he finds that he must pull on it with a force which is constant in magnitude and directed inward toward a fixed point. He reasons that, since the puck is at rest (relative to him) there must be an equal and opposite force acting outward *on the puck*. This somewhat mysterious force he calls *centrifugal* force. On the other hand, the observer on the ground sees that the puck is accelerated toward the center of the merry-go-round and that the rotating observer is applying the needed *centripetal* force. The observer on the ground has no need for the centrifugal force.

Now what about the force that we actually feel when we sit in a car going around a corner? This is the force from the door on us. This force is real enough; but it may be mysterious when we forget that we are moving around a curve. In the reference system of the earth everything is clear. The door of the car must exert a real *centripetal* force on us to cause us to go around the corner instead of straight ahead. In this example the motion in the frame of reference where Newton's law holds is accelerated, and the forces producing the acceleration are real. The car, including the door and us, will not turn the corner unless it is pushed sidewise by the road. The *centrifugal* force, on the other hand, is *not* real. If it were, the net force would be zero, and we would not go around the curve. The centrifugal force is a fiction, something we invent to make it appear that Newton's law of motion applies in the frame of reference of the car.

We see that centrifugal force is really not a force in the sense of a push or a pull exerted by one object on another; rather it is a fictitious or "phony" force that we introduce to correct for the acceleration of our rotating frame of reference.

Centrifugal force is a term which is widely used and poorly understood. If we use it at all, we must remember that it is a device used to make corrections for the dynamical description of motions in a rotating frame of reference. In a frame in which Newton's description is valid, it simply does *not* exist.

14–11 NEWTON'S LAW AND THE ROTATION OF THE EARTH

Now we have raised another problem. We know that the earth spins about its axis once a day relative to the sun or to the fixed stars. Are we in a rotating frame of reference, or are the sun and the fixed stars themselves rotating about the earth? In other words, is Newton's law "really" valid when we use the earth as the reference frame, or is it more accurate in some other frame of reference with respect to which the earth turns? If Newton's law really applies to some frame of reference other than the earth, we should be able to find out by noticing that to explain the observed motion of bodies with respect to the earth we must introduce fictitious forces like centrifugal force. These fictitious forces must be small or they would long since have been apparent in our experiments. Also, we expect them to be small if they correspond to the daily rotation of the earth. To force a mass m to go around daily at the equator requires a centripetal force (Section 14 – 5)

$$4\pi^2 mR/T^2 = (3 \times 10^{-2} \text{ newton/kg})m$$

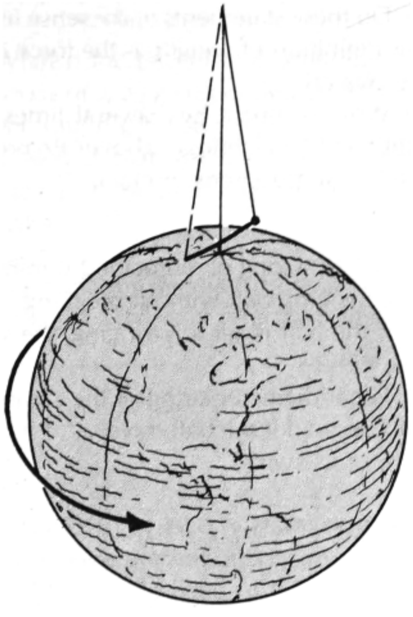
or about $\frac{1}{300}$ of the force of gravity. An experiment to make sure that $\frac{1}{3}$ percent of mg is used to produce this motion must be more precise than the experiments we performed in establishing Newton's law of motion.

The French physicist Foucault performed a famous experiment to demonstrate the fact that the earth rotates on its axis in a frame of reference where Newton's law holds. To perform this experiment we construct a pendulum by hanging a bob on a string, the top end of which is fastened to the ceiling. When set in motion, this pendulum will swing back and forth; and, if the amplitude of the motion is not too large, the bob will perform the simple harmonic motion which we have met in Section 14 – 8.

To start such a bob in motion we might, for example, pull the bob aside from its equilibrium position and release it from rest. It will then oscillate in very nearly straight-line motion. If we are in a proper frame of reference for Newton's law, the motion of the pendulum should continue indefinitely in the plane containing the line of oscillation and the point of suspension of the pendulum (see Fig. 14 – 18). There is no component of force perpendicular to this plane, nor has the pendulum bob an initial velocity component perpendicular to the plane. Therefore Newton's law of motion predicts that, if such a pendulum starts swinging in this plane, it will continue to swing in the same plane indefinitely. This is what will happen in an inertial frame of reference, one in which Newton's law is valid.

When we perform this experiment in the laboratory we find that a pendulum does pretty well; it seems to stay in the plane in which it started and obeys Newton's law of motion in fine shape. But if we wait long enough—hours instead of minutes—we find that slowly but inexorably the plane of the motion rotates relative to its initial orientation. In fact, if we performed this Foucault pendulum experiment at the North Pole we would find that in just 24 hours the plane in which the pendulum swings has rotated completely around through 360 degrees, as seen by us (Fig. 14 – 22). This provides the experimental evidence, obtained from an experiment done on the earth, that the earth is indeed in rotation and that frames of reference rigidly attached to it are rotating frames of reference. It also shows again how little we are likely to notice the effects of the earth's rotation in laboratory experiments. A pendulum with a period of one second oscillates about 10^5 times in a day. If the frictional forces on such a pendulum were so small that it could really swing completely free in the inertial frame, the plane of oscillation would appear to us to rotate through an angle of about 10^{-3} degree in each swing at the North Pole.

These experiments show that the earth is indeed a rotating frame of reference in which Newton's law is not exactly valid. But for all except the most precise experiments we may ignore effects of the earth's rotation. If we want to be as accurate as possible, however, we must use Newton's law in the frame of reference of the fixed stars rather than in that provided by the earth.



14–22 A Foucault pendulum at the North Pole. The pendulum swings in nearly straight-line motion while the earth rotates beneath it. To a man standing on the earth, the plane of the pendulum's motion would seem to rotate.

QUESTIONS

1.* If the gravitational force on an object is 49.0 newtons where the gravitational field is 9.80 nt/kg, what is the object's mass?

2.* The gravitational field at the surface of the moon is one sixth of that at the surface of the earth.
  (a) How much would a 70-kg man weigh on the moon? On the earth?
  (b) What would be his mass on the earth? On the moon?

3. Which two locations would you pick from Table 1 to investigate the change in g with altitude?

4.* What is the magnitude of the gravitational field of the earth at a point where the acceleration of gravity is 9.81 m/sec^2?

5. If different objects had different ratios of their inertial and gravitational masses, would they all accelerate at the same rate in the gravitational field of the earth? Explain.

6.* In Fig. 14 – 2, how long would you make your scale in order to photograph the ball over 30 time intervals, if the ball started from rest at the beginning of the first interval?

7. What is the acceleration of a falling 0.2-kg ball if the gravitational field is 9.80 nt/kg and the air resistance acting on the ball is 0.5 nt?

8. A ball is thrown straight up with a velocity of 15 m/sec.
   (a) How fast will it be going after 1.2 sec?
   (b) How far above the ground will it be at that time?
   (c) How fast will it be going after 2.3 sec?
   (d) How far above the ground will it be at that time?
   (e) What is the ball's acceleration at the top of its rise?

9. A stone is dropped off a cliff of height h. At the same instant a ball is thrown straight up from the base of the cliff with initial velocity v_i.
   (a) Assuming the ball is thrown hard enough, at what time t will stone and ball meet?
   (b) When they meet, is the ball still rising?

10. Two 0.5-kg Dry Ice pucks on a flat table are tied together by a string.
   (a) What is the total gravitational force on these pucks?
   (b) At what rate will they speed up if pulled across the table by a horizontal force equal to the gravitational force found in (a)?
   (c) If one puck remains on the table and the other hangs over the edge, what will be the acceleration of the pucks?

11. An object of mass m is moving on a frictionless inclined plane that makes an angle of 30 degrees with the horizontal.
   (a) What is the net force on the object?
   (b) What is its acceleration down the plane?

12. A stone drops from the edge of a roof and is observed to pass by a window which is 2.0 meters high in 0.10 second. How far is the roof above the top of the window?

13.* How can you tell that the horizontal component of velocity of the ball projected horizontally in Fig. 14 – 4 is constant and its vertical component is not?

14. A baseball is thrown from center field to home plate. If the ball is in the air 3.00 seconds and we neglect air resistance, to what vertical height must it have risen?

15. Plot the trajectory of an object that is thrown into the air with an initial velocity of 10 m/sec at an angle of 45 degrees with the horizontal.

16. Show that the expression $\frac{v^2}{R}$ for centripetal acceleration has the units of acceleration.

17.* What is the centripetal force needed to keep a 3.0-kg object moving in a circle of 2.0 m radius at a speed of 4.0 m/sec?

18.* What is the effect on the speed and direction of motion of an object when a force acts on it in a direction perpendicular to its path?

19.* What is the velocity of the puck in Fig. 14 – 10 when it is at the bottom of the photograph?

20.* What is the acceleration of the puck in Fig. 14 – 10 when it is at the bottom of the photograph?

21. How fast must a plane fly in a loop-the-loop of radius 1.00 km if the pilot experiences no force from either the seat or the safety belt when he is at the top of the loop? In such circumstances the pilot is often said to be "weightless."

22. Some people speak of a pilot coming out of a power dive in a jet airplane as weighing several times his own weight. We sometimes say that the same pilot, falling free before opening his parachute, is weightless.
   (a) Do these statements make sense in the light of our definition of weight as the force of gravity on an object?
   (b) What is meant by "several times his own weight" and "weightless" above? Be prepared to discuss these questions in class.

23. A string 5.0 m long of diameter 2.0 mm just supports a hanging ball without breaking.
   (a) If the ball is set to swinging, the string will break. Why?
   (b) What diameter string of the same material should be used if the ball travels 7.0 m/sec at the bottom of its swing?

24. An electron (mass $= 0.91 \times 10^{-30}$ kg) under the action of a magnetic force moves in a circle of 2.0 cm radius at a speed of 3.0×10^6 m/sec. At what speed will a proton (mass $= 1.6 \times 10^{-27}$ kg) move in a circle of the same radius if it is acted upon by the same force?

25. A body moving with a speed v is acted upon by a force that always acts perpendicular to the mo-

tion of the body but increases steadily in magnitude.

(a) Draw a sketch of the trajectory.

(b) Does the speed of the object increase, decrease, or remain unchanged?

26.* How must the force on an object vary with the displacement if simple harmonic motion is to take place?

27.* At what points on the circle in Fig. 14 – 17 will *P* have the same velocity as *N*?

28.* If the force constant of a spring is 8 newtons/meter, what suspended mass will give a period of 1 sec?

29.* What must be the length of a simple pendulum if it is to have a period of one second?

30. A block of mass *m* rests on a horizontal platform. The platform is driven vertically in simple harmonic motion with an amplitude of 0.098 meter. When at the top of its path, the block just leaves the surface of the platform. (This means that at this point its acceleration is 9.8 m/sec^2 downward.)

(a) What is the period of the simple harmonic motion?

(b) When the block is at the bottom point of its path, what is its acceleration?

(c) What is the force exerted by the platform on the block at this bottom point?

31. (a) What is the period of a pendulum consisting of a mass of 2.0 kg suspended on a light cord 2.4 meters long if $g = 9.8$ newtons/kg?

(b) Notice that we can use a pendulum to measure *g*. If the period of this pendulum is 3.0 sec, what is *g*?

32. A 2-kg ball is suspended from a spring. When disturbed in a vertical direction, the ball moves up and down in simple harmonic motion at a frequency of 4 cycles per second.

(a) What is the period of the motion?

(b) What is the upward force exerted on the ball by the spring when the ball is at the midpoint of its up-and-down path?

(c) How much did the spring stretch when the ball was first attached to its end (before the oscillatory motion was started)?

33. A Dry Ice puck on a horizontal surface is attached to four identical springs as shown in Fig. 14 – 23. Points *A* and *B* are pinned when the puck is at rest in the midposition. The puck is displaced toward *A* by an amount *d* and then released. While it is oscillating between two springs with the amplitude *d* and a period *T*, the pins *A* and *B* are removed as the puck passes through the midposition. This allows the puck to vibrate under the action of all four springs.

(a) What is the relation of the force constant *k′* of two identical springs in series compared with the force constant *k* of one of the springs?

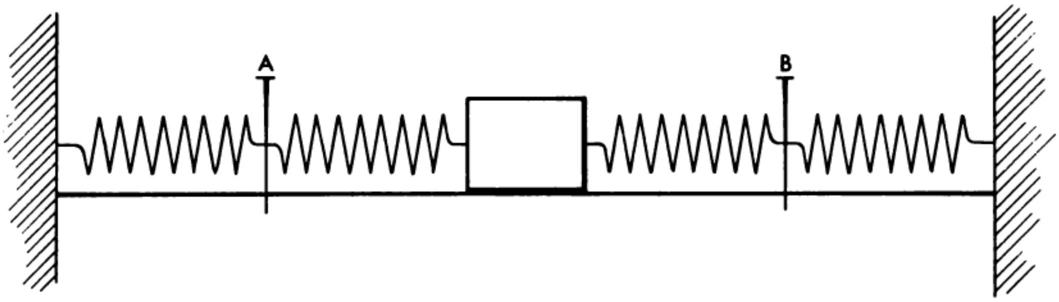
(b) What is the new period of vibration?

(c) Has the speed at the midposition been changed?

(d) What is the new amplitude of the puck?

34. A ball is thrown vertically upward with a velocity of 24 meters/sec from a railroad flatcar moving horizontally with a velocity of 4 meters/sec. Describe the path of the ball as seen by an observer (a) on the flatcar, and (b) by the observer on the ground nearby.

35. While a bus is in motion along a level, straight road, we roll a marble from one side to the other across the floor of the bus. Its path is a straight line relative to the bus. Later we roll it again, and this time the path is a parabola which bends toward the front of the bus (concave toward the front of the bus). Describe the motion of the bus in each case. Be prepared to discuss in class.



14–23 For Question 33.

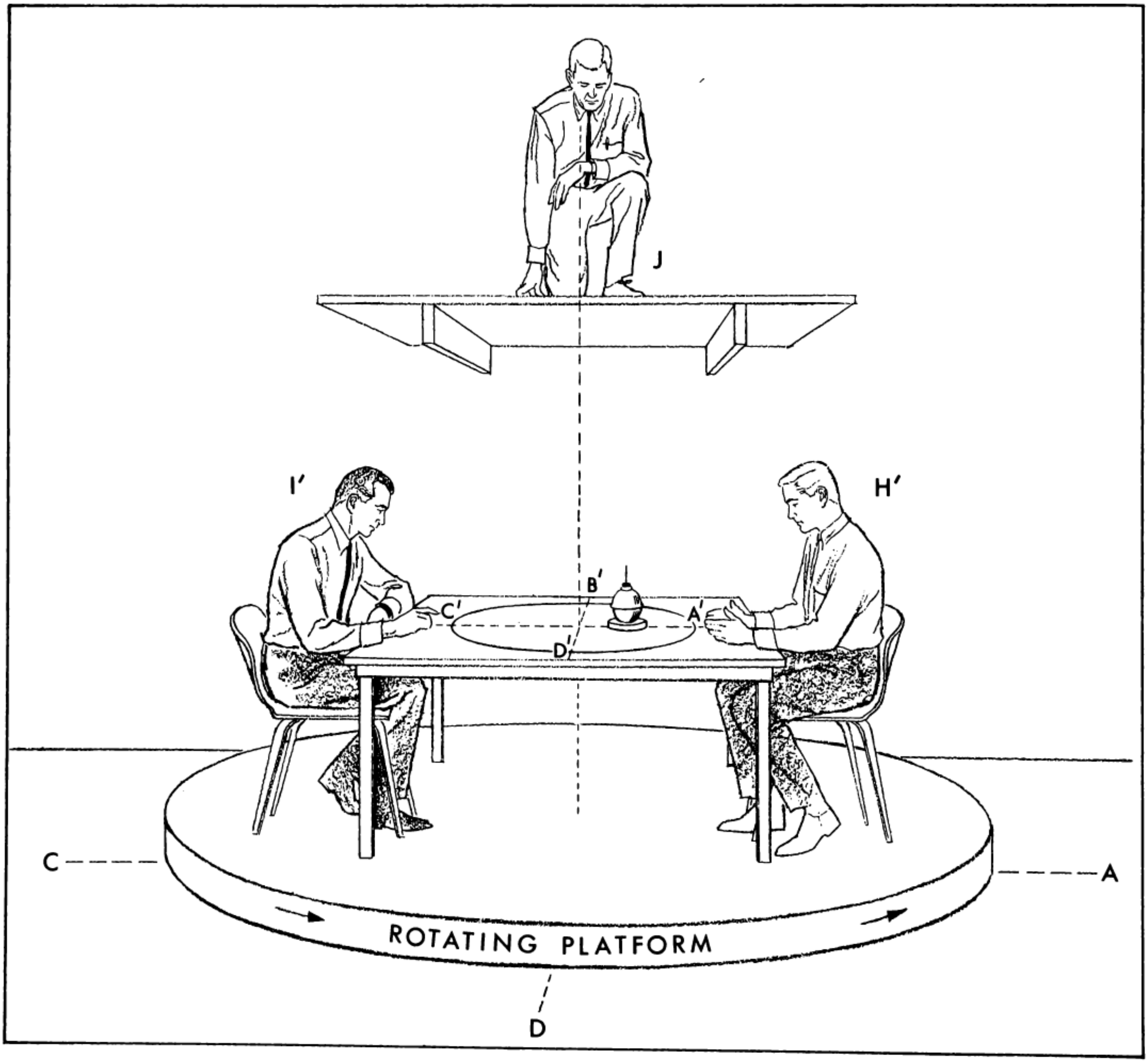
36. For some experiments in relative motion, a smooth, level table, shown in Fig. 14 – 24, is centered on a platform which rotates uniformly at a rate of one revolution in 12 seconds. Two perpendicular lines are drawn through the center of the table, intersecting a circle of 1.20-meter radius at the points *A′*, *C′*, *B′*, and *D′*. Two men, *H′* and *I′*, sit on the platform at opposite ends of the line *A′C′*. A third man, *J*, is above the table so that he can observe the motion of a Dry Ice puck in a stationary frame of reference. He has four marks on the floor, forming two perpendicular reference lines *AC* and *BD* through the center of the table. (In the diagram, *B* is behind the table.)

(a) *H′* holds the puck in his hand at *A′*. What is its velocity in *J*'s frame of reference at the instant *A′* crosses the line *AC*?

(b) As *H′* passes *A* he gives the puck a sudden push so that it travels along line *AC* with a speed of 0.40 m/sec. Construct a vector diagram to show the velocity that *H′* gave the puck in *J*'s frame of reference.

(c) Construct a diagram which shows the positions of the puck and of *A′* at 1, 2, 3, 4, 5, and 6 seconds in *J*'s frame of reference.

(d) Make use of the diagram constructed for (c) to construct a diagram which will show the motion of the puck on the table, as seen by *H′* and *I′*. (Hint: Construct this moving-frame dia-



14–24 For Question 36.

gram on tracing paper which can be laid over the diagram of the fixed-frame motion. Plot the position of the puck for each 15° rotation of the moving frame. These points can be connected to give a picture of the motion in the moving frame.)

37. With the same situation as in Question 36:

(a) Suppose that *H′* launches the puck as he passes *A* so that *I′* will catch the puck as he passes *D*. Construct a diagram which shows the motion of the puck in *J*'s frame of reference. What is the speed of the puck in this frame of reference?

(b) Make use of the diagram constructed for (a) to construct a diagram which will show the motion of the puck on the table as seen by *H′* and *I′*.

(c) Construct a vector diagram to show the velocity that *H′* gave the puck so that it could travel as in (a).

38. With the same situation as in Question 36:

(a) With what speed and in what direction can *H′* launch the puck as he passes *A* so that, as *J* sees it, the puck remains at *A*?

(b) What is the motion of the puck as seen by *H′* and *I′*?

(c) The puck has a mass of 0.50 kg. After *H′* has let go of it, what force acts on it according to *J*? What force acts on it according to *H′* and *I′* to make it move as it does in their frame of reference?

39.* A long-range gun is located in the Northern Hemisphere. You wish to hit a point due north of it. In what direction should you aim the gun?

15

Universal Gravitation and the Solar System

THERE ARE few people who do not recognize the Big Dipper in the night sky. The most striking fact about the heavens is that the constellations, groups of stars such as the Big Dipper, maintain unaltered shapes. They move just as though they were tacked on the inside of a large rotating sphere and we were at the center of this sphere watching (Fig. 15 – 1). Against this background of "fixed stars" the sun and moon move smoothly, as if they were attached to other spheres rotating at different rates about the earth. According to this view, the earth, large and immobile, is located at the center of a universe made of celestial matter which revolves about it. Such a universe is called *geocentric* (earth-centered).

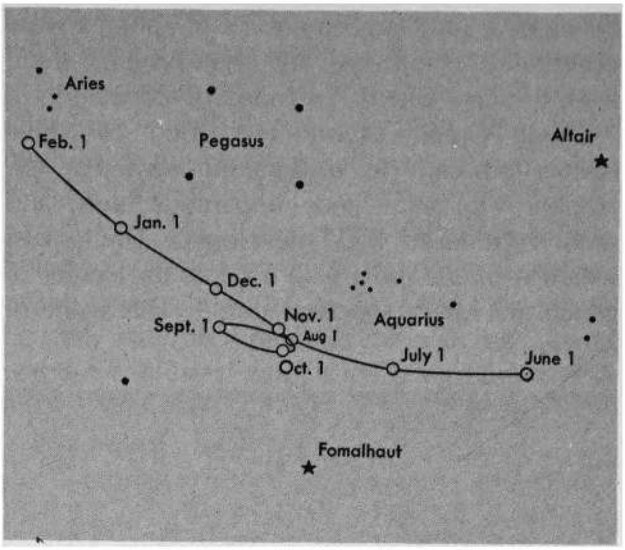
Seven heavenly bodies which appeared to move among the fixed stars were known to ancient man. The sun and moon, Mercury, Venus, Mars, Jupiter, and Saturn were called planets from the Greek word meaning "wanderer." With the exception of the motion of the sun and moon, the motions of these bodies appear irregular when viewed over long periods of time (Fig. 15 – 2). Their erratic motion focused the attention of ancient men on the planets. They were brighter than the stars; and because their brightness changed, their distances from the earth seemed to change. They became associated with various human endeavors and emotions (Venus with love, Mars with war, etc.), as though they formed an intermediary between the immutable perfection of the stars and the restless imperfection of the earth. Later, the astrologers saw in the planetary positions indications of the future course of the lives of human beings.

Finding a reasonable explanation for the peculiar observed motion of the planets was a major concern of the ancient astronomers. It is related that Plato, the Greek philosopher (427 – 347 B.C.), set the following problems for his students: The stars, as we can see, move in perfect circular paths around the earth, but the planets seem to trace irregular figures. What are the combinations of perfect circular paths along which the planets really move? The form of this question reveals the fact that the circle was believed to be the most perfect of all curves and hence alone worthy of describing celestial motions. The efforts of all astronomers for many centuries were devoted at least partly to answering this question.

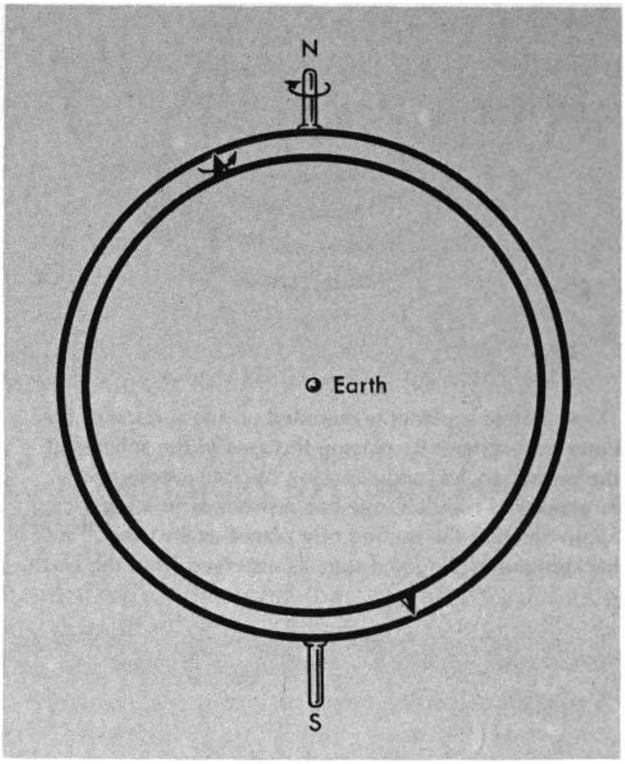
15–1 A one-hour time exposure taken with the camera pointed at the North Star. The circular arcs show the apparent motion of the stars. It was this circular motion that led the Greeks to visualize the stars as attached to a sphere which turned about the earth. (Yerkes Observatory Photograph.)

15–1 EARLY PLANETARY SYSTEMS

Plato's pupil Eudoxus tried to represent planetary motions by a collection of moving spheres, each with its center at the earth. Each planet was attached to the surface of a sphere which rotated uniformly about an axis attached to two opposite points on the surface of a larger sphere (Fig. 15 – 3). While the inner sphere spun uniformly on its axis, the axis itself was carried around by a uniform motion of the outer sphere. In fact, the axis of the outer sphere might itself be attached to the surface of a still larger sphere; in this way the number of spheres could be extended to represent more complex motions. Finally, the whole system spun around inside the celestial sphere which carried the fixed stars. With a sufficient number of spheres moving around inside of other spheres, Eudoxus obtained a good approximation to the motion of a planet. His successors improved the accuracy of his model by using still more spheres. In the course of history many planetary systems depending on the motions of the spheres were



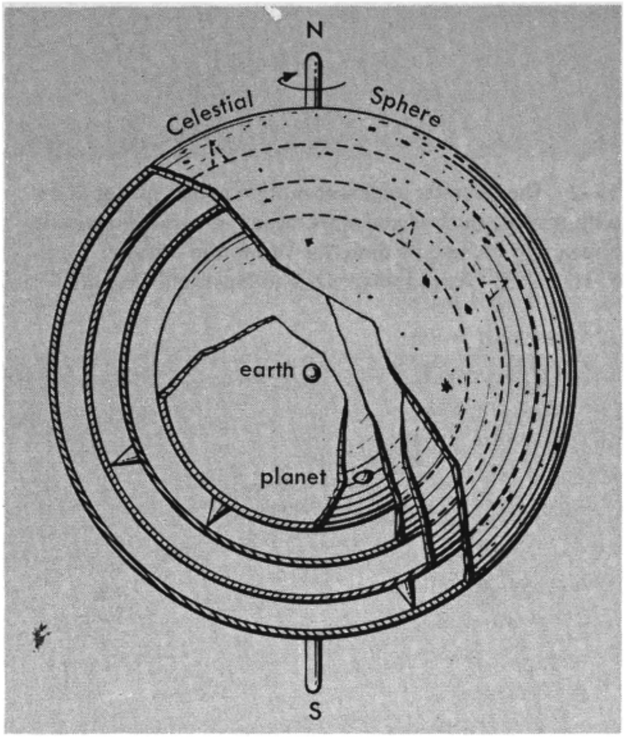
15–2 The peculiar apparent motion of the planet Mars with respect to the fixed stars. Mars, at various intervals, appears to reverse its direction of motion. (From: R. H. Baker, "Astronomy," D. Van Nostrand Co., Inc.)



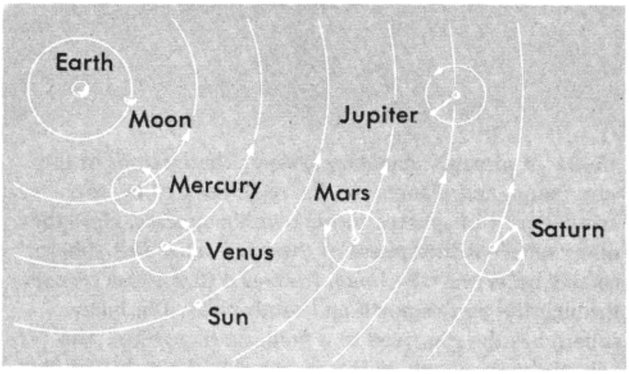
15–3a Eudoxus's planetary system. The motions of the sun, moon, and planets can be approximated by an arrangement of spheres turning within spheres. Here the outer sphere is the sphere of the fixed stars. This sphere rotates once every 24 hours from east to west on an axis through the earth's north and south poles. The inner sphere has the sun fixed to a point on its surface. The axis of the inner sphere is connected to the outer sphere at two points. The inner sphere completes a rotation once a year.

developed, and large numbers of spheres were eventually employed. In one system, there were thirteen spheres for Mercury alone.

Other Greek astronomers tried to solve Plato's problem in a different way. For example, Apollonius and Hipparchus (third and second centuries B.C.) developed a system in which a planet moves on a circle the center of which moves on another circle. The work of these early Greek astronomers led to the system of Claudius Ptolemy of Alexandria in the second century A.D. His system of circles moving on other circles reproduced the observed motions of the planets reasonably accurately (Fig. 15 – 4). But his curves describing the planetary orbits were so complicated that there were many complaints from those who studied them. Alphonso X, King of Castile in 1200, was prompted to say that, had he been consulted at the creation, he would have made the world on a simpler and better plan. A simplified Ptolemaic orbit of a planet is shown by the heavy line in Fig. 15 – 5; the circular motions from which this orbit was supposed to have originated are also shown.



15–3b Here a planet is mounted on the surface of the innermost sphere. By placing the axes of the spheres at the proper angles, and choosing suitable speeds and directions of rotation, one can reproduce to a good approximation the motion of a planet against the background of the fixed stars as observed from the earth.



15–4 A simplified diagram of the Ptolemaic system of planetary motions.

15–2 COPERNICUS' PLANETARY SYSTEM

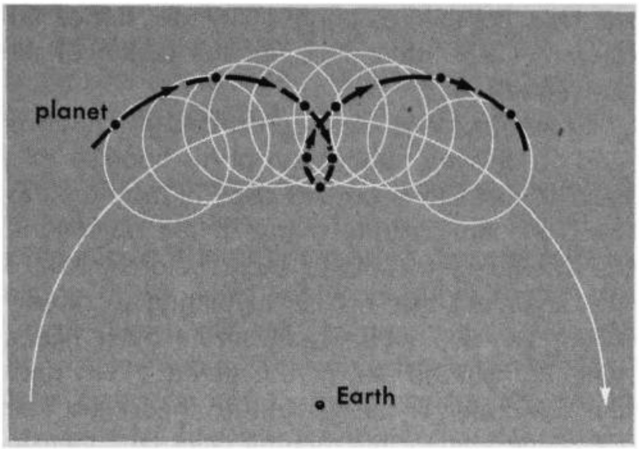
The Polish astronomer Nicolaus Copernicus (born in 1473) felt that the Ptolemaic system was too complex. The simplicity of uniform circular motion desired by Plato had been buried in complicated constructions. The truth, Copernicus thought, must be simpler. He therefore set out to give a simpler answer to Plato's problem by choosing a different center for the system of circles.

Like others before him, Copernicus realized that the motion of the fixed stars might be explained by assuming that the earth rotates.* Our situation with respect to the heavens is then much like that of a passenger in an airplane. As the airplane turns in flight above a big city, the streets and avenues seem to go around. So as the earth rotates, the stars appear to move.

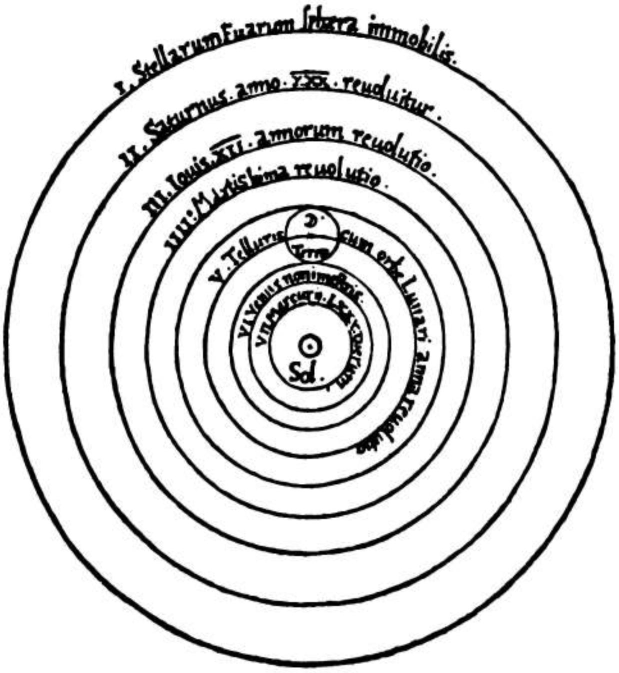
Once he imagined that the earth rotates daily, Copernicus found that the orbits of the planets could be greatly simplified by choosing the sun rather than the earth as the center of the planetary system. Then the earth would be neither at the center of the universe nor at rest. Perhaps it was a planet, and revolved along with the other planets about the sun. In Fig. 15 – 6 we show the simple orbits followed by the earth and

---

*Although the main current of Greek and medieval thought embodied a geocentric theory, Heracleides (about 370 B.C.) believed that the earth rotated on its axis, and Aristarchus (third century B.C.) thought that the earth moved around the sun.



15–5 The motion of a single planet in the Ptolemaic system. The planet was considered to revolve in a small circle whose center moved in a circular orbit about the earth.


I. Stellarum Fixarum ſphera immobilis.
II. Saturnus anno XXX. reuoluitur.
III. Iouis XII. annorum reuolutio.
IIII. Martis bima reuolutio.
V. Telluris cum orbe Lunari anno reuolutio
Terra
VI. Venus nono mense.
VII. Mercurius LXXX. dierum.
Sol.

15–6 The orbits of the planets according to the planetary system of Copernicus.

the planets as they move around the sun according to Copernicus.

Copernicus justified his proposal that the earth moved by saying, "And although it seemed to me an absurd opinion, yet, because I knew that others before me had been granted the liberty of supposing whatever circles they chose in order to demonstrate the observations concerning the celestial bodies, I considered that I too might well be allowed to try whether sounder demonstrations of the revolutions of the heavenly bodies might be discovered by supposing some motion of the earth. . . . I found after much and long observation, that if the motions of the other planets were added to the motions of the earth, . . . not only did the apparent behavior of the others follow from this, but the system so connects the orders and sizes of the planets and their orbits, and of the whole heaven, that no single feature can be altered without confusion among the other parts and in all the universe. For this reason, therefore, . . . have I followed this system."

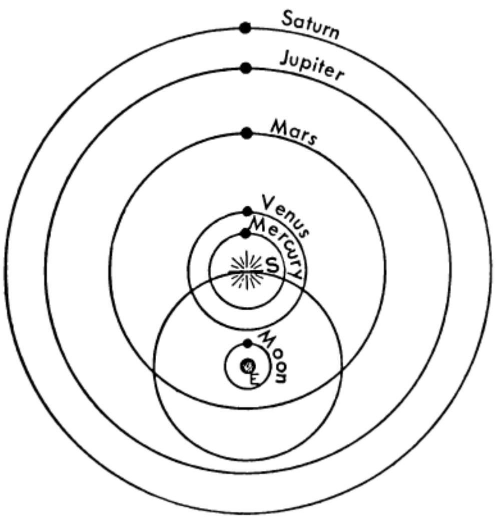
Copernicus' system also included a large, immobile sphere on which the fixed stars were located. About it he said: "The first and highest of all the spheres is the sphere of the fixed stars. It encloses all the other spheres and is itself self-contained; it is immobile; it is certainly the portion of the universe with reference to which the movement and positions of all the other heavenly bodies must be considered. If some people are yet of the opinion that this sphere moves, we are of a contrary mind; . . ." Then, describing the planetary spheres and their periods of rotation in which the earth appears as one of six planets while the moon is clearly designated as a satellite of the earth, he concluded: "In the midst of all, the sun reposes unmoving. Who, indeed, in this most beautiful temple would place the light-giver in any other part than that whence it can illumine all the other parts? . . ."

15–3 OBJECTIONS TO COPERNICUS' THEORY

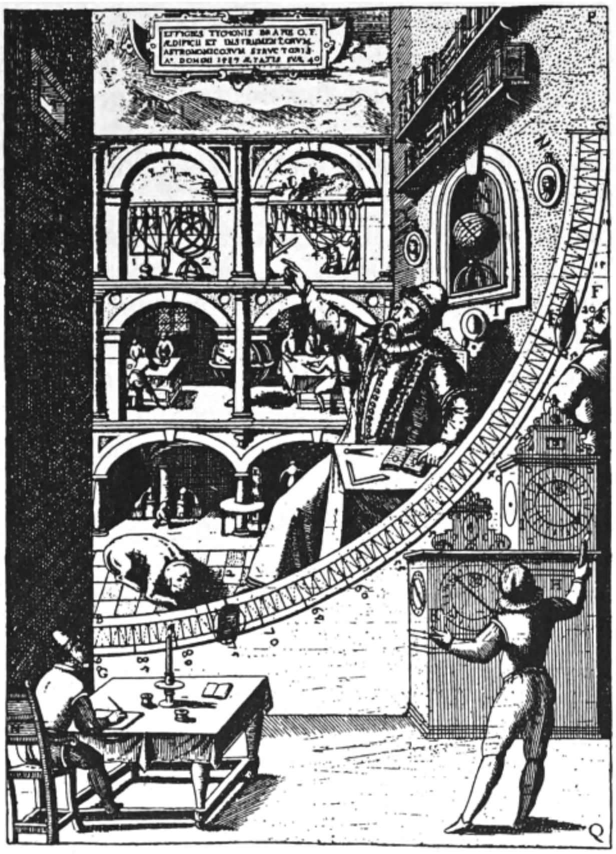
It must be understood that in order to achieve his simplified orbits Copernicus was obliged to discard the entire picture of the universe that had been developed from the time of Aristotle. The question as to whether or not the earth moved was a very serious one. All of medieval cosmology and physics was based on the idea that the earth was at rest at the center of the universe. In part this belief was based on man's inner conviction that his earth must be at the center of things. But in addition there seemed to be good evidence of the earth's special position. For one thing, if the earth moves, what is pushing it and why is this motion not felt? Or, to take another example, why do stones fall toward the earth if it is not the center of the universe?

Copernicus expected a great deal of criticism, and he delayed publication of his book so long that he first saw a printed copy of it on the day he died. Anticipating many of the objections, he attempted to answer them beforehand. To the argument that his earth rotating so rapidly about its own axis would surely burst like a wheel driven too fast, he countered: "Why does the developer of the geocentric theory not fear the same fate for his rotating celestial sphere—so much faster because so much larger?" To the argument that birds in flight would be left behind by the rapidly moving earth, Copernicus answered that the atmosphere is dragged along with the earth.

In fact, there were many arguments and counterarguments. The Copernican theory was denounced as "false and altogether opposed to the Holy Scriptures." Martin Luther branded Copernicus a fool and a heretic. The argument over this bold new concept of the universe raged for more than 100 years before the notion that the earth could move was generally accepted.



15–7a Tycho Brahe's geocentric system.



15–7b Tycho Brahe's mural quadrant. A brass arc of over 6-foot radius was mounted on a wall and equipped with movable sights. An observer (upper right) sighted on a star through the window in the wall at the left. Other assistants noted the time and recorded the angle of the observation. The sector of wall enclosed by the arc is covered with a mural painting of Tycho and other aspects of his work.

15–4 TYCHO BRAHE

Tycho Brahe, the Danish astronomer born in 1546, could not accept the Copernican system, despite its simplicity. He contributed instead an improved geocentric system in which the sun goes around the earth and the other planets go around the sun [Fig. 15 – 7 (a)]. Also, to test astronomical models, he set out to make a really accurate map of the positions of the fixed stars and to determine the apparent positions of the planets as seen from the earth over a long period of time. He began his observations with an instrument consisting of a pair of joined sticks, one leg to be pointed at a fixed star, the other at a planet. In this way he could measure their angular separation. Later he constructed great sextants and compasses with which he made wonderfully careful observations [Fig. 15 – 7 (b)]. He catalogued the positions of 777 stars so accurately that his observations are still used, and his measurements of the planetary angular positions over a period of twenty years contain no error larger than 4/60 of a degree. This angle is about the same as the angle subtended by the head of a pin held at arm's length from your eye.

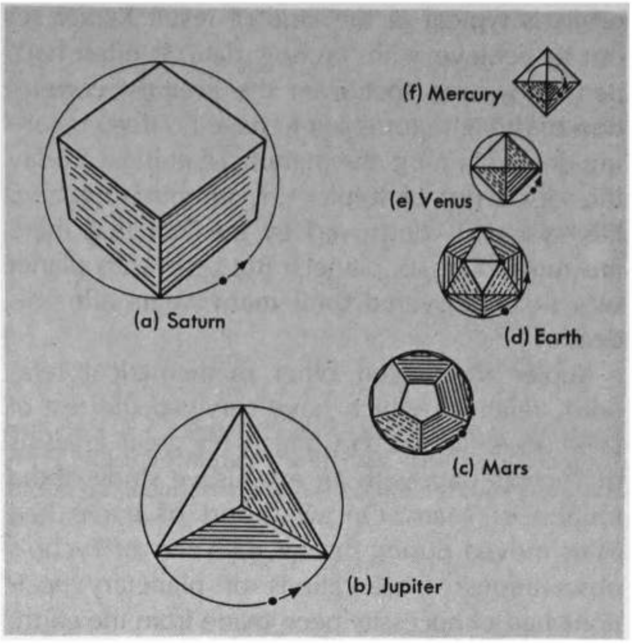
Tycho's observations of planetary positions were far more accurate than those available to Copernicus. They soon showed that the Copernican orbits were only roughly correct. A search then began for a more accurate description of the orbits. This objective was reached, after Tycho's death, by the German astronomer Kepler, who had worked in Tycho's laboratory during the last eighteen months of Tycho's life.

15–5 KEPLER

Johannes Kepler, born in 1571, was a striking contrast to Tycho Brahe. Tycho possessed tremendous mechanical ability and skill but relatively little interest in mathematics. Kepler was clumsy as an experimenter but was fascinated by the power of mathematics. He was akin to the ancient Greeks in his reverence for the power of numbers and was intrigued by puzzles concerning number and size.

After Kepler had learned the elements of astronomy, he became obsessed with the problem of finding a numerical scheme underlying the planetary system. He wrote, "I brooded with the whole energy of my mind on this subject." He devoted his life to the analysis of the tables of planetary positions which Tycho had left him. In tackling the problem of translating the observations of Tycho Brahe into mathematical descriptions of planetary motions, Kepler acted like any scientist today who attempts to explain experimental findings in terms of simple mathematical laws rather than mere tables of numbers. With mathematical laws we can not only reproduce observed data, but we can predict the results of observations not yet made. Furthermore, mathematical laws are easier to remember and to communicate than mere tables of numbers.

In his first book Kepler described his attempts to understand why there were precisely six planets in the solar system. He established a connection between the six orbits and the five regular geometrical solids* (Fig. 15 – 8). From this construction he obtained ratios of radii agreeing fairly well with the values known for the planetary orbits.



15–8 Kepler's law of planetary orbits was based on the five regular solids. According to the law, a sphere, with a radius equal to that of the orbit of Saturn, is circumscribed about a cube (a). A sphere inscribed within this cube has a radius equal to the radius of the orbit of Jupiter.

In (b) the sphere of the orbit of Jupiter is shown with a tetrahedron inscribed. A sphere inscribed within the tetrahedron gives the radius of the orbit of Mars.

In (c) the sphere for Mars has a dodecahedron incribed. A sphere inscribed within that gives the orbit of the earth (d).

We may continue this process of alternate inscribing of spheres and regular solids, using the icosahedron (20-sided) and the octahedron. These will give us the orbits of Venus (e) and of Mercury (f), which is on a sphere inscribed inside the octahedron. Kepler considered the five solids as bridging the intervals between planetary orbits. Since there are only five regular solids, Kepler believed there could be only six planets.

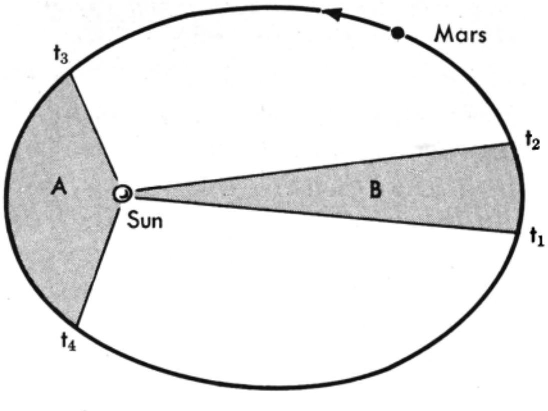
Kepler was ecstatic. He wrote, "The intense pleasure I have received from this discovery can never be told in words. I regard no longer the time wasted; I tired of no labor; I shunned no toil of reckoning, days and nights spent in calculation, until I could see whether my hypothesis would agree with the orbits of Copernicus, or whether my joy was to vanish into air."

The relation between the radii of planetary orbits is typical of the kind of result Kepler set out to achieve with Tycho's data. It often happens, however, that even the prettiest correlation of the data turns out to have no deep meaning in explaining the nature of things. Today this discovery of Kepler's is all but forgotten. His system is destroyed by the fact that there are more than six planets. But a seventh planet was not discovered until many years after his death.

---

*By a regular solid body we mean a symmetrical body with equal flat faces. Only five kinds of regular solid bodies can be constructed.

Kepler discovered other mathematical relations, relations which have survived the test of later observations. He began his great analysis of Tycho's data with an exhaustive study of the motion of Mars. On what sort of curve had Mars moved during the twenty years of Tycho's observations? Observations of planetary positions had of necessity been made from the earth. Would Mars move in a simple curve if one imagined the earth to be at rest, or if one imagined the earth to be in motion as Copernicus believed? Kepler adopted the Copernican idea that the earth spun about its axis while moving in an orbit about the sun. Following tradition, he first tried a system of circles moving on circles to obtain possible orbits. He made innumerable attempts, each involving long and laborious calculations. He had to translate each of Tycho's measurements of the angle between Mars and the fixed stars into a position of the planet in space with respect to a fixed sun about which the earth itself was moving.



15–9 Kepler's law of equal areas. Mars travels along its orbit with varying speed, moving fastest when it is closest to the sun. Kepler found that for equal time intervals, $t_2 - t_1 = t_4 - t_3$, the areas swept out by a line between the sum and the planet were equal (area B = area A). In the drawing, the elongation of the ellipse has been exaggerated to illustrate the equal-area law more clearly.

After about seventy trials using the "eccentric circle" type of orbit, Kepler found one scheme which agreed fairly well with the facts. Then, to his dismay, he found that this curve, when continued beyond the range of the data he had used, disagreed with other observations that Tycho had made of Mars' position.

This disagreement between Tycho's data and Kepler's calculations was about $\frac{8}{60}$ of a degree. (This is the angle through which the second hand on a watch moves in about 0.02 second.) Might not Tycho have been wrong by this small amount? Could not the cold of a winter's night have numbed his fingers or blurred his observations? Kepler knew Tycho's methods and the painstaking care he took. Tycho could never have been wrong even by such a small amount. Thus, on the basis of Tycho's data, Kepler rejected the curves he had constructed. What a tribute this was to the experimental skill of Tycho Brahe!

Saying that "upon this eight minutes [he] would yet build a theory of the universe," Kepler began again. Discarding the ancient and cherished belief in uniform motion, he considered possible changes of the speed of a planet as it moved in its orbit about the sun. He now made his first great discovery. He found that a line from the sun to a planet sweeps out equal areas in equal times. This has come to be known as Kepler's second law (Fig. 15 – 9).

After the discovery of his second law, Kepler finally abandoned his attempts to construct planetary motions out of combinations of uniform circular motions and began to try various ovals as possible orbits. After many more laborious calculations, he finally achieved one of his most important results—his so-called first law. Each planet, he found, moves in an elliptical orbit with the sun at one focus. Imagine Kepler's delight. After years of effort, he had finally found a simple curve which described the motions of planets.

Kepler then set to work to find a connection between the size of a planet's orbit and its period, the time of one revolution of the planet about the sun. After many trials he found the precise relation for which he was searching: for all the planets the ratio of the cube of the radius

TABLE 1 *Kepler's Third Law*

PLANET	RADIUS R OF ORBIT OF PLANET (A.U.)	PERIOD T (DAYS)	R^3/T^2 $[(\text{A.U.})^3/(\text{DAY})^2]$	MODERN VALUES R^3/T^2 $(\text{M}^3/\text{SEC}^2)$
Mercury	0.389	87.77	7.64×10^{-6}	3.354×10^{18}
Venus	0.724	224.70	7.52	3.352
Earth	1.000	365.25	7.50	3.354
Mars	1.524	686.98	7.50	3.354
Jupiter	5.200	4,332.62	7.490	3.355
Saturn	9.510	10,759.20	7.430	3.353

The values of the orbits and periods in this table are those used by Kepler. In Kepler's day, the radii were known only in terms of the radius of the earth's orbit. The radius of the earth's orbit is called an astronomical unit (A.U.) of length. The nearly constant values of R^3/T^2 illustrate Kepler's third law. The last column is based on accurate modern measurements of the orbits and periods.

TABLE 2 *The Solar System*

OBJECT	MASS (KILOGRAMS)	RADIUS (METERS)	PERIOD OF ROTATION (SECONDS)	MEAN RADIUS OF ORBIT (METERS)	PERIOD OF REVOLUTION (SECONDS)
Sun	1.98×10^{30}	6.95×10^{8}	2.14×10^{6}	——	——
Mercury	3.28×10^{23}	2.57×10^{6}	7.60×10^{6}	5.79×10^{10}	7.60×10^{6}
Venus	4.83×10^{24}	6.31×10^{6}	2.6×10^{6}(?)	1.08×10^{11}	1.94×10^{7}
Earth	5.98×10^{24}	6.38×10^{6}	8.61×10^{4}	1.49×10^{11}	3.16×10^{7}
Mars	6.37×10^{23}	3.43×10^{6}	8.85×10^{4}	2.28×10^{11}	5.94×10^{7}
Jupiter	1.90×10^{27}	7.18×10^{7}	3.54×10^{4}	7.78×10^{11}	3.74×10^{8}
Saturn	5.67×10^{26}	6.03×10^{7}	3.60×10^{4}	1.43×10^{12}	9.30×10^{8}
Uranus	8.80×10^{25}	2.67×10^{7}	3.88×10^{4}	2.87×10^{12}	2.66×10^{9}
Neptune	1.03×10^{26}	2.48×10^{7}	5.69×10^{4}	4.50×10^{12}	5.20×10^{9}
Pluto	?	?	?	5.9×10^{12}	7.82×10^{9}
Moon	7.34×10^{22}	1.74×10^{6}	2.36×10^{6}	3.8×10^{8}	2.36×10^{6}

of the orbit to the square of the period is the same.* Once he happened upon this ratio, the regularity was striking. (See Table 1.) The constancy of the ratio R^3/T^2 is called Kepler's third law.

With this triumph Kepler wrote, ". . . what sixteen years ago I urged as a thing to be sought . . . that for which I joined Tycho Brahe . . . at last I have brought to light and recognize its truth beyond my fondest expectations. . . . The die is cast, the book is written to be read now or by posterity. I care not which—it may well wait a century for a reader as God has waited six thousand years for an observer."

Kepler had carried astronomy through a momentous advance. He had translated the magnificent tables of data of Tycho Brahe into a simple and comprehensive system of curves and rules. Kepler's system earned him the title "Legislator of the Heavens."

---

*The radius R of an orbit is defined by taking one half the sum of the shortest and longest distances between the sun and the planet. Because the planetary orbits are not very different from circles, the distance from the sun to any point on a planetary orbit will do for most purposes.

Here are statements of Kepler's three laws:

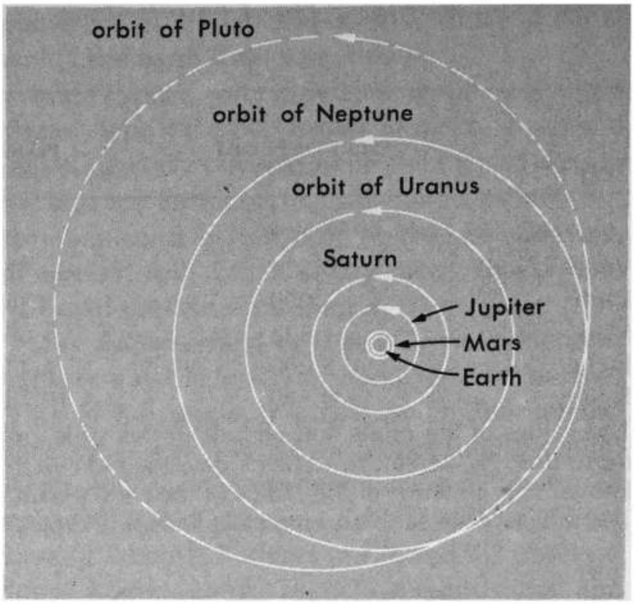
I. Each planet moves in an elliptical path with the sun at one focus.

II. The line joining the sun and the planet sweeps out equal areas in equal times.

III. The ratio $\frac{R^3}{T^2}$ is the same for all planets. If this constant ratio is called K, the third law may be written

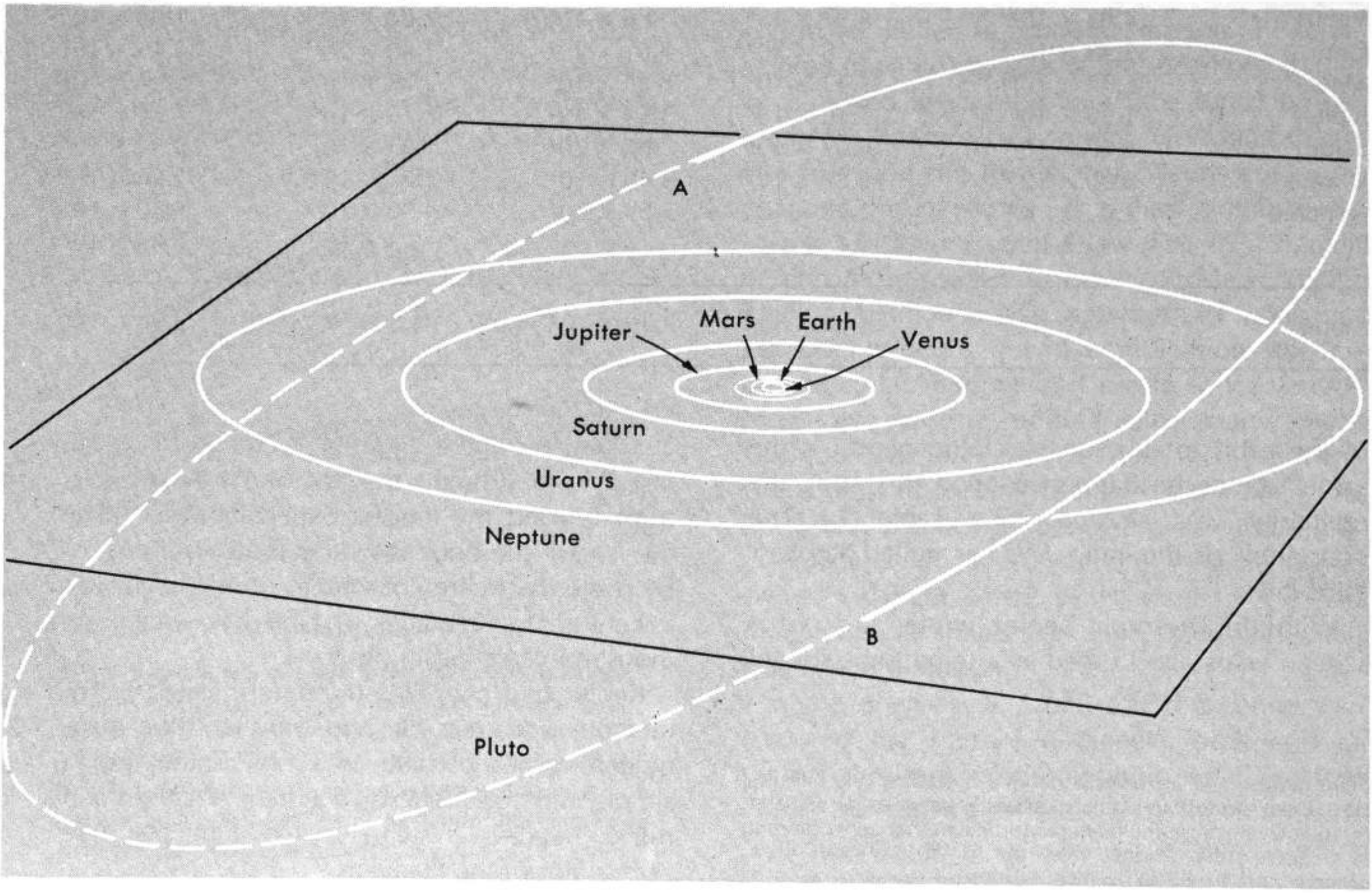
$$\frac{R^3}{T^2} = K$$

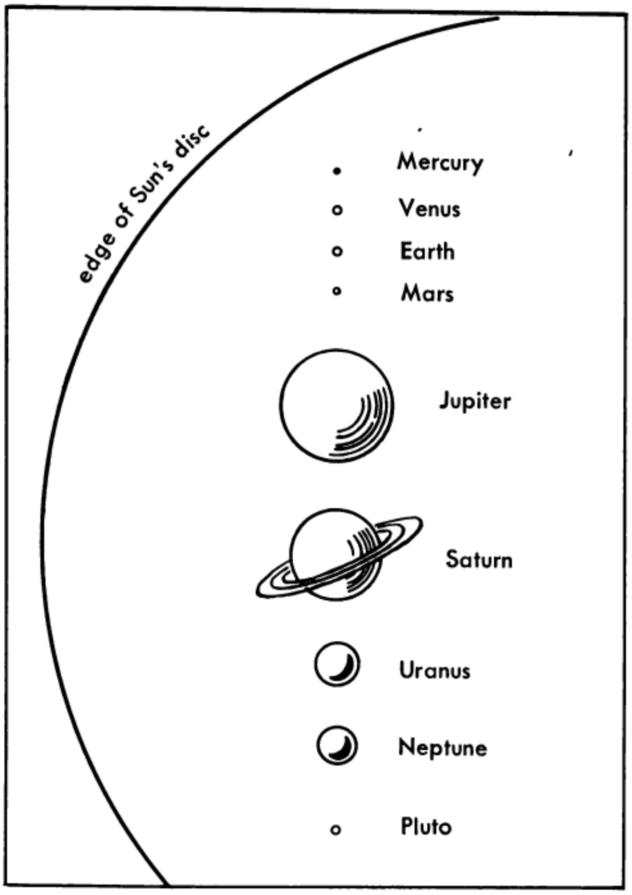
Kepler's three laws give a more accurate representation of the planetary orbits than had been obtained from either the Ptolemaic or the Copernican system with all their complexity of circles moving on circles.



15-10 **Approximate orbits of the principal planets. (The orbits of Mercury and Venus are too small to show on this drawing.) They are very nearly circles, except for Pluto, and are in almost the same plane. Only careful measurements show them to be ellipses. The orbits of Pluto and Mercury are the most elliptical, and the plane of Pluto's orbit makes an angle of about 17° with the planes of the other orbits. In the figure, the portion of Pluto's orbit below the plane of the paper is shown by a dashed line.**

15-11 **The orbits of the known planets as seen from an angle. Notice that all but the outermost planet, Pluto, revolve about the sun in approximately the same plane. In this figure the orbit of Mercury is too small to show.**





15–12 The approximate sizes of the planets compared to the sun. The total mass of all the planets equals only 1.34 × 10^{-3} of the sun's mass.

TABLE 3 *A Scale Model of the Solar System*

It is impossible to show in a small drawing both the relative sizes and the distances of the planets on the same scale. The following table gives some idea of the relative sizes and distances in terms of common objects. To get the true dimensions of the solar system, every dimension should be multiplied by 4.4×10^9.

OBJECT IN THE SOLAR SYSTEM	OBJECT IN THE MODEL	DISTANCE FROM "SUN"
Sun	basketball	
Mercury	half a pinhead	13 m
Venus	appleseed	25 m
Earth	appleseed	34 m
Mars	small appleseed	52 m
Jupiter	golf ball	180 m
Saturn	Ping-Pong Ball	320 m
Uranus	a marble	0.65 km
Neptune	a marble	1.0 km
Pluto	?	1.3 km
Nearest star	basketball	8×10^3 km

15–6 KINEMATIC DESCRIPTION AND THE DYNAMICAL PROBLEM

Kepler's laws are the kinematics of the planetary system. They give a simple, accurate description of the motions of the planets, but they do not provide an explanation of the motions in terms of forces. Ptolemy's description of planetary motion is also kinematics. What is the essential difference in the two kinematic descriptions?

Both descriptions are reasonably accurate. Each allows us to predict where we shall see a particular planet at a given time. The difference lies in the point of view. Kepler's description appears simpler to us. He followed Copernicus in choosing the fixed stars as a framework with respect to which the planetary motions are specified, and like Copernicus he used the sun as the origin from which to measure the planetary positions. It is simpler to describe planetary motions using the sun as the origin than it is to describe the motions as seen from the earth, and we can conclude from this simple description what the planetary motions must look like when viewed from earth.

On the other hand, there is nothing wrong in describing the motion directly, using the earth as the origin. It is merely difficult and confusing because the motions we see look complicated and irregular. Convenience—not principle—determines the frame of reference and origin we use in kinematics.

The situation is like that of a man on the ground watching the motion of a point on the edge of the wheel of a car (Chapter 12). He sees the point move on a cycloidal path with a speed that varies periodically from zero to a maximum value. On the other hand, a man who describes the motion with respect to the axle of the wheel finds that the point moves at constant speed around a circle. Both men are right; and if we take account of the motion of the axle with respect to the ground, the two descriptions are equivalent. In just the same way, geocentric and heliocentric descriptions of the planetary systems can be equivalent, as are the systems of Copernicus and of Tycho Brahe (Figs. 15 – 6 and

15 – 7); and which we adopt is a matter of convenience. For navigation, for example, the earth-centered view is preferable—we don't care how simple the motions would appear from the sun. We want to know where we are when we see the planets in certain directions at a certain time. Consequently, the geocentric language is used in celestial navigation of ships and airplanes today.

When we wish to explain planetary motions, however, the situation is different. In the first place we should certainly suspect that the dynamical explanation will be easier to arrive at and to understand, the simpler the description of the motion. In the second place, we already know that Newton's law of motion applies only in certain frames of reference. Even the uniform motion that goes on in the absence of a net force on an object will not look uniform to an observer who spins around like a top. In this case we know that complicated fictitious forces will appear, and then we cannot find a simple dynamical explanation. Consequently, in making the transition from kinematics to dynamics, it is important to find a preferred frame of reference where fictitious forces do not confuse us.

To explain planetary motion dynamically, we need to choose an appropriate frame of reference. Can we choose that frame so that the earth is at rest in it? To this question the answer seems to be "no." In any such frame, the motions of the planets imply irregular forces and no dynamical explanation has been found. We can only retire to the Aristotelian view that planets are different from other matter and behave according to their own special laws of motion.

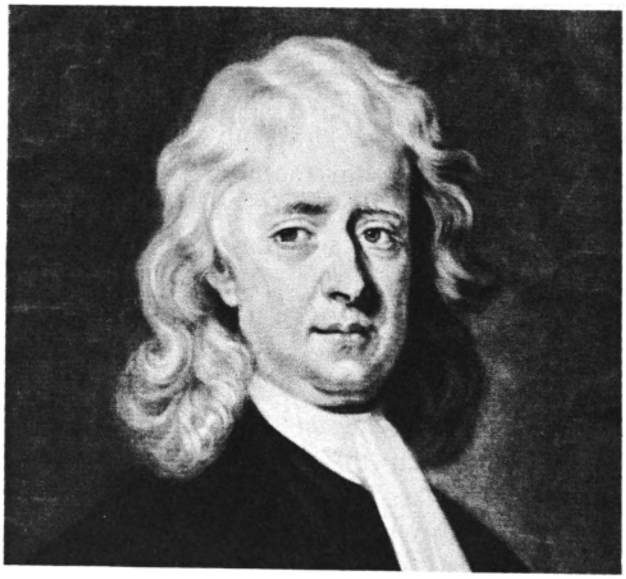
The Ptolemaic system was appropriate to this Aristotelian view. It also fitted with a geocentric dynamics for earthly objects. It led to the idea that the orbits of the planets should appear simplest when viewed with the earth as center (and not around some other point). The greater simplicity of the planetary orbits in Kepler's heliocentric description thus undermines the whole Aristotelian picture.

In the heliocentric system, on the other hand, the earth becomes a planet like other planets. Then there is no reason left to have a special geocentric dynamics. Instead we can look again for a single dynamics that includes the motions of objects on the earth and of all the planets including our own. Indeed the heliocentric view provided the clue from which we have constructed a dynamical explanation of planetary motion.

In the rest of this chapter we shall see how the heliocentric system fits with the new dynamics of Galileo and Newton. There we shall follow in the footsteps of Newton. With the heliocentric description and a frame of reference tied to the fixed stars, we shall find that a simple law of force between chunks of matter leads to the observed planetary motions. It explains them on the basis of the same dynamics that applies here on earth. Since the time when this law of gravitational force was postulated by Newton, it has also been experimentally tested for small pieces of matter here on earth. It is valid here as well as through astronomical distances. Thus the search for a dynamics in which motions here and in the heavens are of just the same kind was successful. On a geocentric picture no such unified explanation is available.

15–7 NEWTON

Isaac Newton was born in 1642, the year Galileo died. He brought together the discoveries of Copernicus, Kepler, Galileo, and others in astronomy and in dynamics. To these he added his own findings and fused them into a structure that still stands today, one of the greatest achievements of science. So profound and clear was his understanding, that he was able to apply the laws of motion successfully to an astonish-



15–13 Sir Isaac Newton.

ing number of phenomena, from the movement of the planets to the rise and fall of the tides.

Between the time of Kepler and that of Newton a great change had taken place in scientific thought. After Galileo's work the feeling grew that there were universal laws governing the motion of bodies and that these laws might apply to motion in the heavens as well as on earth. The scientific discussions in the Royal Society of London often centered about the question, "What sort of force does the sun exert on the planets which causes the planets to move according to the laws which Kepler has discovered?" With Kepler's laws as a guide Newton answered this question. He created a planetary dynamics which was so successful that for many years scientists complained that nothing was left to be done.

Newton's first effort to understand the motion of celestial bodies was directed toward a study of the motion of the moon. Newton knew that if no force acted on the moon it would move in a straight line with constant speed. However, as viewed from the earth, the moon follows a nearly circular path. Consequently, there must be an acceleration toward the earth and a force which produces it. He stated:

"Nor could the moon without some such force be retained in its orbit. If this force was too small, it would not sufficiently turn the moon out of a rectilinear course; if it was too great it would turn it too much and draw the moon from its orbit toward the earth."

What is the force that causes the moon to move about the earth? Newton said that the answer came to him while he was sitting in a garden. He was thinking about this problem when an apple fell to the ground; the force which the earth exerted on the apple might, he thought, also be exerted on the moon. The moon might be a falling body.

In Chapter 14 we calculated the acceleration of the moon toward the earth and found it to be about 2.7×10^{-3} m/sec^2, nowhere near the value of 9.8 m/sec^2 which is the acceleration of a falling body at the earth's surface. Newton carried out essentially the same calculation. At first he did not have available a very accurate value for the radius of the moon's orbit. He did know, however, that it was about sixty times the radius of the earth; and, using a rough value for the earth's radius, he could obtain the radius of the moon's orbit and calculate the moon's acceleration. When he found out how small the acceleration of the moon is, Newton must have asked himself questions like these: Why is the acceleration of a falling body so much greater than that of the moon? Does the force with which the earth attracts a body decrease as the body gets farther and farther away? If so, what is the exact relation between the force and the distance of separation?

Newton had assumed that the earth pulled on the moon in the same general way that it pulled on the falling apple. If this assumption was to be retained, any postulated law of force would have to explain the acceleration g of a body at the earth's surface and the much smaller value of the acceleration of the moon. Newton explained many years later that he was led to the correct force law by working backward from Kepler's third law. He temporarily left the forces that were exerted by the earth and considered instead the forces exerted by the sun on the planets, the centripetal forces which kept the planets moving in their orbits. Newton wanted to know how the force on a planet varied with the radius of the planet's orbit. We shall now see how this force may be calculated.

One of Kepler's triumphs was his description of planetary orbits as ellipses. But the planetary orbits are nearly circular, and for simplicity we shall approximate them as circles with the sun at their common center. Let us consider a planet moving around the sun with a period T in a circular orbit of radius R. As we learned in Section 14 – 5, the centripetal acceleration of a planet or any object moving uniformly around a circle is $a = \frac{4\pi^2 R}{T^2}$. Therefore the centripetal force on the planet must be

$$F = ma = \frac{m4\pi^2 R}{T^2}$$

where m is the mass of the planet. This is the force that acts on the planet.

In order to eliminate the period T and express the force as a function of R and m alone, Newton used Kepler's third law $\frac{R^3}{T^2} = K$, or $T^2 = \frac{R^3}{K}$. On substituting $\frac{R^3}{K}$ for T^2 in the equation $F = \frac{m4\pi^2 R}{T^2}$, we find the force on the planet is

$$F = 4\pi^2 K \frac{m}{R^2}$$

The force is proportional to the mass of the planet and inversely proportional to the square of the distance from the sun (Fig. 15 – 14).

Eventually Newton was able to show that any body moving under the action of this force must move in an elliptical orbit with the sun at one focus and that the line connecting the sun with the body will sweep out equal areas in equal times. Furthermore, we know from the method by which we found the force that Kepler's third law will follow from it. The whole system of planetary motion described by Kepler's law, therefore, follows from this law of force and Newton's law of motion.*

15–8 UNIVERSAL GRAVITATION

Notice that the factor $(4\pi^2K)$ in the law of force between the sun and a planet enters the equation from the law of periods. It applies to any planet of any mass on any orbit around the sun. Therefore, $(4\pi^2K)$ depends only on the properties of the sun; it measures the strength of the sun as the source of the force of attraction.

The force between the sun and a mass m is

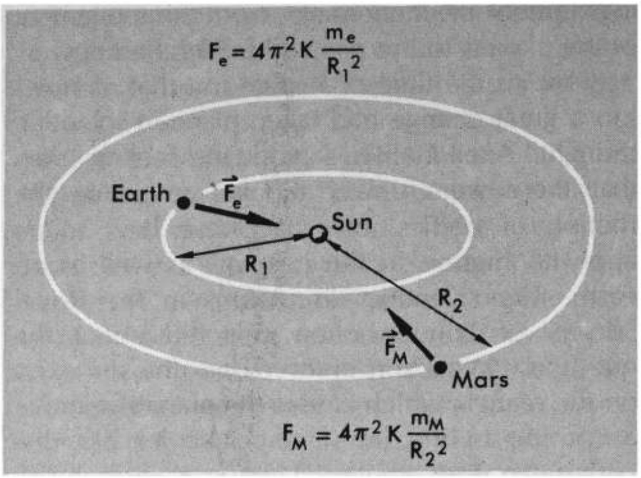
$$F = \frac{(4\pi^2K_s)m}{R^2}$$

where $4\pi^2K_s$ refers to the sun and R is the distance between the sun and the mass m. Perhaps the force between the earth and a mass m is

$$F = \frac{(4\pi^2K_e)m}{R^2}$$

where $(4\pi^2K_e)$ is the strength of the earth as a source of gravitational attraction and R is now the distance between the earth and the mass m. With such ideas Newton returned to the problem of the moon's motion about the earth. Measuring R from the center of the earth, the value of the gravitational field strength g (the acceleration of a mass falling at the earth's surface) is then

$$g = \frac{F}{m} = \frac{(4\pi^2K_e)}{R_e^2}$$



$F_e = 4\pi^2K\frac{m_e}{R_1^2}$
Earth
$\vec{F}_e$
Sun
R_1
R_2
$\vec{F}_M$
Mars
$F_M = 4\pi^2K\frac{m_M}{R_2^2}$

15–14 The gravitational force exerted by the sun on a planet is proportional to the mass of the planet and inversely proportional to the square of its distance from the sun.

where R_e is the radius of the earth, the distance between the center of the earth and a mass m at its surface. Also, the value of the gravitational field strength at the distance of the moon, which is the acceleration of the moon toward the earth, is

$$a_m = \frac{(4\pi^2K_e)}{R_m^2}$$

where R_m is the distance from the center of the earth to the center of the moon. On dividing this equation by the last one we get

$$\frac{a_m}{g} = \frac{R_e^2}{R_m^2} \quad \text{or} \quad a_m = g\,\frac{R_e^2}{R_m^2}$$

Since Newton Knew that $\frac{R_e}{R_m}$ is about 1/60 and g is 9.8 m/sec², he found that $a_m \approx 2.7 \times 10^{-3}$ m/sec². This is about the same value of a_m that he obtained from the radius and period of the moon (Section 15 – 7).

Now Newton had obtained the acceleration of the moon in two different ways: from R_m and the period of the moon's motion without any reference to the inverse-square law of force, and from g at the earth's surface by using the inverse-square law in the ratio $(R_e/R_m)^2$. The approximate agreement of the values he obtained gave him support for the idea that the force between earth and moon was of the same kind as that between the sun and the planets. They were both gravitational forces like the force on the falling apple.

Newton probably did not show that all of Kepler's laws follow from the law of gravitational force until several years after his original discoveries. But he discovered the law of force and applied it to the problem of the moon when he was twenty-four years old. Of this period Newton later wrote: "And the same year I began to think of gravity extending to ye orb of the Moon, and . . . from Kepler's Rule (third law) . . . I deduced that the forces which keep the Planets in their Orbs must (vary) reciprocally as the square of their distances from the centers about which they revolve: and thereby compared the force requisite to keep the Moon in her Orb with the force of gravity at the surface of the earth, and found them to answer pretty nearly. All this was in the two plague years of 1665 and 1666, for in those days I was in the prime of my age for invention, and minded Mathematicks and Philosophy more than at any time since."

---

*Huygens and Hooke also used Kepler's third law and Newton's law of motion to infer that F is proportional to $1/R^2$, but they did not show that Kepler's other laws then followed. Newton provided the law of motion, found the law of force, and also showed that it resulted in Kepler's description of planetary motions.

Newton certainly suspected that the inverse-square law of attraction applied not only to the sun and planets, not only to the earth and moon, but also to any two chunks of matter. This suspicion leads immediately to the question: What property of a body determines its gravitational attraction for other masses? What property of the earth determines how large $4\pi^2K_e$ is for the earth? What determines $4\pi^2K_s$ for the sun? Perhaps $4\pi^2K$ depends on a new property of a body; but if gravitational attraction is a property of all bodies, it is reasonable to suppose that $4\pi^2K$ depends on the quantity of matter in the body. The simplest assumption is that $4\pi^2K$ is proportional to the mass of the body. Then $4\pi^2K_e = Gm_e$ for the earth, and $4\pi^2K_s = Gm_s$ for the sun, where G is the proportionality factor between $4\pi^2K$ and m for any body.

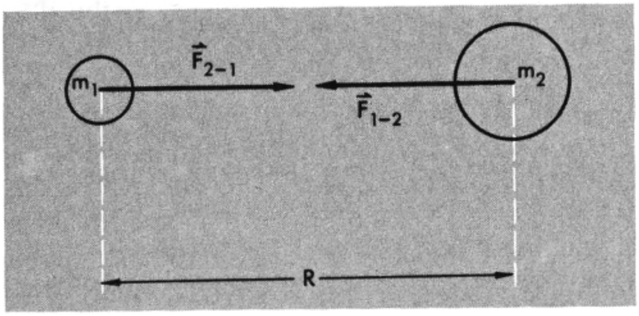
Newton made this assumption. With it, the magnitude of the gravitational force of attraction that a body of mass m_1 exerts on a body of mass m_2 at a distance R becomes

$$F_{1-2} = (4\pi^2K_1)\frac{m_2}{R^2} = Gm_1\frac{m_2}{R^2}$$

Furthermore, because every mass attracts every other piece of matter gravitationally, the mass m_2 also exerts a gravitational force on m_1. Since $4\pi^2K_2 = Gm_2$, the magnitude of this force of attraction exerted by m_2 on m_1 is

$$F_{2-1} = (4\pi^2K_2)\frac{m_1}{R^2} = Gm_2\frac{m_1}{R^2}$$

Therefore, the two forces are equal in magnitude, although they are opposite in direction (Fig. 15 – 15). The expression $F = G\frac{m_1m_2}{R^2}$ for the magnitude of the attraction summarizes Newton's law of universal gravitation: any two bodies attract each other with forces proportional to the mass of each and inversely proportional to the square of the distance between them. The universal gravitational constant, G, does not depend on what objects are considered, where they are, or what their state of motion may be.

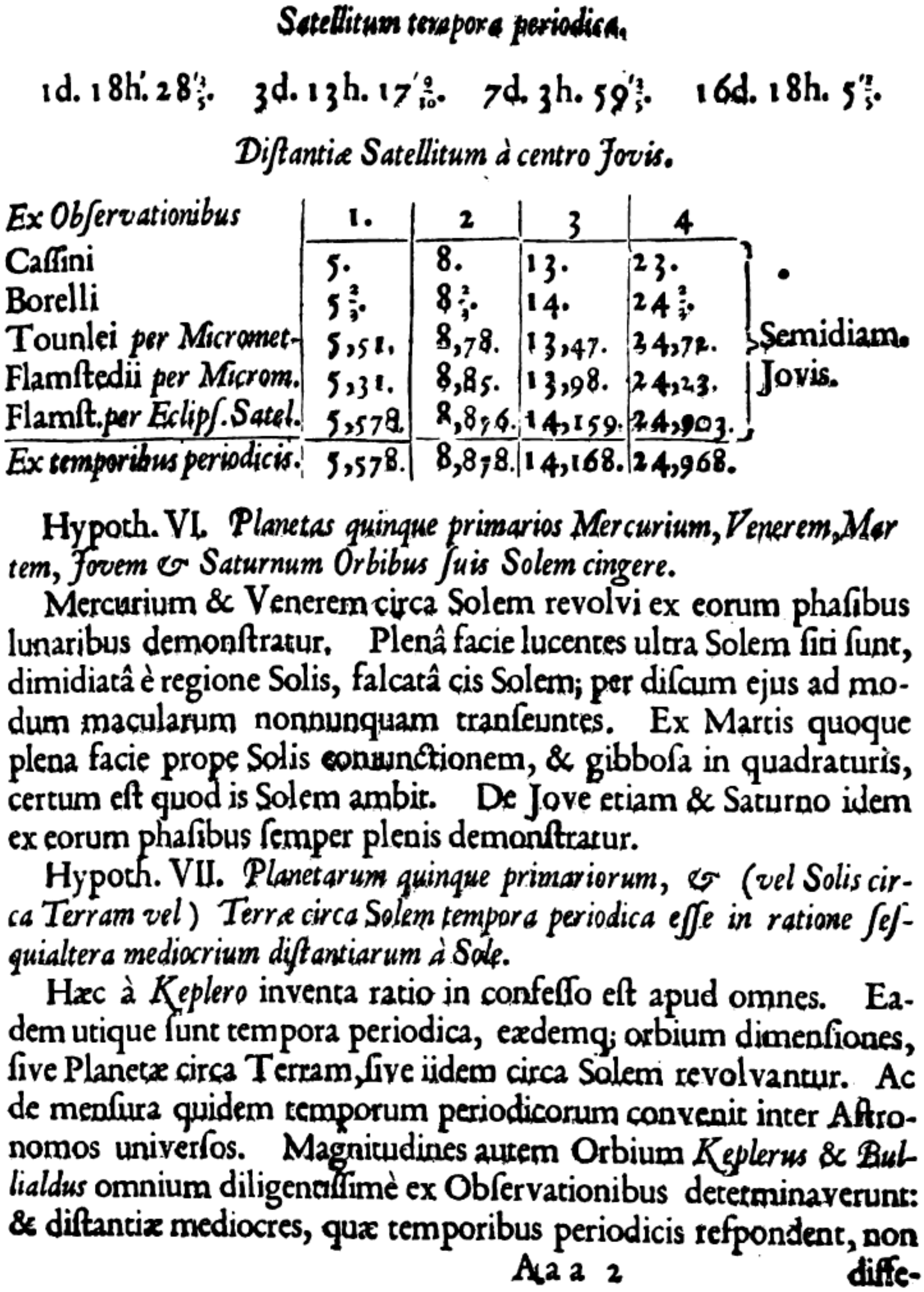


15–15 The force of gravitation exerted by m_1 on m_2 is equal and opposite to the force exerted by m_2 on m_1.

We do not know in detail how Newton arrived at the law of universal gravitation. In addition to the reason that we have given to make the law plausible, a number of other considerations may have suggested the same result. For example, as we shall see in the next chapter, Newton eventually assumed that the forces of interaction between two bodies are always equal and opposite, and he may already have had this idea in mind when he formulated the law of universal gravitation. Whatever the steps by which Newton was led to the law of universal gravitation, the law eventually stands or falls by the consistency between the predictions we can make from it and the actual behavior of the universe. On the basis of this law Newton was able to take great strides in building up a theoretical system of the universe.

15–9 SOME OF NEWTON'S LATER ACCOMPLISHMENTS

Newton applied the law of universal gravitation to a wide variety of problems. As we mentioned previously, he derived all three of Kepler's empirical laws from the law of universal



Satellitum tempora periodica.

1d. 18h. 28'$\frac{3}{5}$. 3d. 13h. 17'$\frac{9}{10}$. 7d. 3h. 59'$\frac{3}{5}$. 16d. 18h. 5'$\frac{1}{5}$.

Diſtantiæ Satellitum à centro Jovis.

Ex Obſervationibus	1.	2	3	4	
Caſſini	5.	8.	13.	23.	Semidiam. Jovis.
Borelli	5$\frac{2}{3}$.	8$\frac{2}{3}$.	14.	24$\frac{2}{3}$.	
Tounlei *per Micromet.*	5,51.	8,78.	13,47.	24,72.	
Flamſtedii *per Microm.*	5,31.	8,85.	13,98.	24,23.	
Flamſt. *per Eclipſ. Satel.*	5,578.	8,876.	14,159.	24,903.	
Ex temporibus periodicis.	5,578.	8,878.	14,168.	24,968.	

Hypoth. VI. *Planetas quinque primarios Mercurium, Venerem, Martem, Jovem & Saturnum Orbibus ſuis Solem cingere.*

Mercurium & Venerem circa Solem revolvi ex eorum phaſibus lunaribus demonſtratur. Plenâ facie lucentes ultra Solem ſiti ſunt, dimidiatâ è regione Solis, falcatâ cis Solem; per diſcum ejus ad modum macularum nonnunquam tranſeuntes. Ex Martis quoque plena facie prope Solis conjunctionem, & gibboſa in quadraturis, certum eſt quod is Solem ambit. De Jove etiam & Saturno idem ex eorum phaſibus ſemper plenis demonſtratur.

Hypoth. VII. *Planetarum quinque primariorum, & (vel Solis circa Terram vel) Terræ circa Solem tempora periodica eſſe in ratione ſeſquialtera mediocrium diſtantiarum à Sole.*

Hæc à *Keplero* inventa ratio in confeſſo eſt apud omnes. Eadem utique ſunt tempora periodica, eædemq; orbium dimenſiones, ſive Planetæ circa Terram, ſive iidem circa Solem revolvantur. Ac de menſura quidem temporum periodicorum convenit inter Aſtronomos univerſos. Magnitudines autem Orbium *Keplerus* & *Bullialdus* omnium diligentiſſimè ex Obſervationibus determinaverunt: & diſtantiæ mediocres, quæ temporibus periodicis reſpondent, non

Aaa 2 diffe-

15–16 A page from Newton's "Philosophiae Naturalis Principia Mathematica" (London, 1687). The periods of revolution of the four largest moons of Jupiter are given at the top of the page. The table gives the radii of the orbits of the moons as measured by different observers. The bottom line of the table gives the radii as calculated from the periods using Kepler's third law.

gravitation. He then examined the tides and explained them on the basis of the gravitational force exerted by the moon, both on the earth and on the oceans. He began to analyze the small irregularities (perturbations) of planetary orbits. These small deviations of the planets from their predicted elliptical paths can be explained by the small gravitational interactions among the planets themselves. Not only is the earth attracted by the sun, but also in varying degree it is attracted by each of the other planets. These attractions are relatively small because of the small masses of the planets in comparison with the mass of the sun; but their effects can be observed and they are predicted correctly by Newton's theory.

Later this perturbation theory led to the discovery of a new planet. In the nineteenth century seven planets were known. Of these, the original six behaved well; but the seventh, Uranus, which had been discovered by Herschel in 1781, did not act quite as expected. When the perturbations of its orbit by the other planets had been computed, the result did not agree with the details of the observed motion. The astronomers Adams and Leverrier arrived independently at the conclusion that there must be another yet undiscovered planet (still farther away from the sun but close enough to influence the motion of Uranus); and on September 23, 1846, the astronomer Galle found the new planet where Leverrier had told him to look. This new planet was named Neptune.

Among the many other problems to which Newton applied the law of universal gravitation, one is of particular interest to us. It concerns the calculation of the acceleration of the moon from the inverse-square law of force and the value of g at the surface of the earth (Section 15 – 8). When he originally performed this calculation, Newton used the distances R_e and R_m from the center of the earth. Although the center of the earth is a natural place from which to measure R, Newton was not sure that it was the right place. Because he suspected that the inverse-square law of attraction applied to any two pieces of matter, Newton thought that the gravitational attraction of the earth for an object should be the resultant of the forces attracting it toward each piece of matter in the earth.

The different pieces of matter in the earth are located at various distances from an object at the earth's surface. Do they, all acting together, produce the same force on this object that they would produce if all were concentrated at the center of the earth? Does the force fall off as $1/R^2$ even close to the surface of the earth? Before he would be satisfied, Newton had to solve the mathematical problem of adding up all the forces arising from all the pieces of matter of which the earth is made. He had to prove that this vector sum gave the inverse-square law of force outside the surface of the earth.

Now we can solve this problem by applying an elegant mathematical theorem; but in Newton's time this theorem and its basis were unknown. Newton himself invented the mathematics (now called the calculus) necessary to solve this and similar problems. When he obtained the answer it turned out that his original assumption was correct. When the forces from each piece of matter fall off as the square of the distance, spherical bodies do attract one another as though the entire mass of each were concentrated at its center. Newton was delighted.

15–10 LABORATORY TESTS OF THE LAW OF UNIVERSAL GRAVITATION

The direct way to see if Newton's law of universal gravitation is consistent with the behavior of all matter is to measure the gravitational forces between chunks of matter in the laboratory. We should find the gravitational attraction between two masses and measure the force on each; we should use objects of various materials to see that the mass alone determines the attraction. If the law were found to be consistent with our measurements, we would then naturally evaluate the universal constant of proportionality G.

Such experiments are difficult. Even when two stones are placed close together, they do not attract each other noticeably. With a crude estimate of G we can see why. According to the law of universal gravitation, the gravitational force on a mass m at the surface of the earth is

$$F = G\,\frac{m_e m}{R_e^{\,2}}$$

The gravitational field strength g is therefore

$$g = \frac{F}{m} = G\,\frac{m_e}{R_e^{\,2}}$$

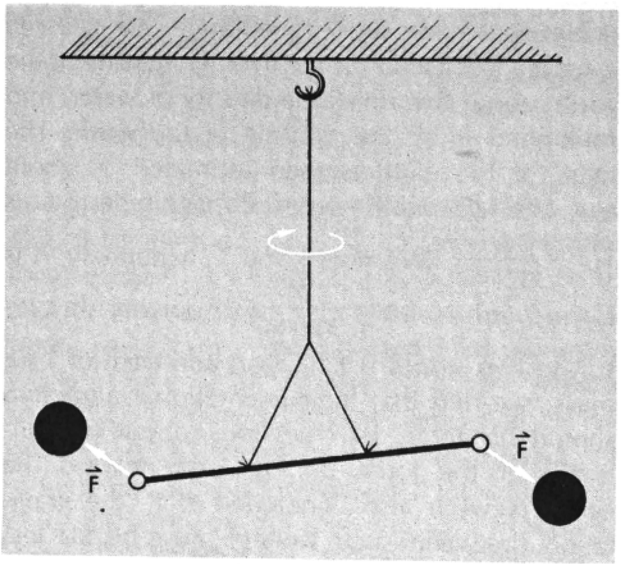
In this equation we know g; it is 9.8 m/sec². We know the radius of the earth $R_e = 6.38 \times 10^6$ meters; and Newton knew an approximate value for it. To determine G, therefore, we need only to estimate the mass of the earth.

Newton made such an estimate. He guessed a reasonable value for the average density of the earth, about five times the density of water, and multiplied it by the volume of the earth. The mass of the earth is then estimated as about 6×10^{24} kg; and the order of magnitude of G is $10^{-10}\,\frac{\text{m}^3}{\text{kg-sec}^2}$. We now know G accurately. It is $G = 0.667 \times 10^{-10}\,\frac{\text{m}^3}{\text{kg-sec}^2}$. Applying this result to two stones 0.1 m apart and each of 1 kg mass, we find that their gravitational attraction should be about 10^{-8} newtons, about one billionth of the force pulling them toward the earth. Newton also concluded that "the gravitation [between such stones] must be far less than to fall under the observations of our senses." He therefore directed his attention to the calculations about the gravitational interactions of the planets and their satellites which we mentioned in the last section.

One hundred years later, in 1798, Lord Cavendish succeeded in measuring the gravitational interaction of laboratory-sized objects. The apparatus he used is diagramed in Fig. 15 – 17. Two small spheres were mounted on opposite ends of a light rod 2 meters long. This rod hung horizontally with its center below a fine vertical wire from which it was suspended. At the ends of the rod and on the side of the case which contained the apparatus, Cavendish mounted ivory rulers with which to measure the position of the rod. When Cavendish placed two large masses near the small spheres at the ends of the rod, the small spheres were attracted toward the big masses and the wire suspension twisted.

Cavendish recorded the position of the rod with the suspension twisted when the big masses were located as shown in Fig. 15 – 17. Then he moved each of the large masses to the symmetrical position on the other side of the little spheres. The gravitational attraction twisted the suspension in the opposite direction, and he measured the new position. By measuring the changes in position when known forces were applied to the small balls, Cavendish could find the strength of the gravitational forces between the small balls and the large masses.*

Actually Cavendish performed numerous experiments. He had to evaluate possible extraneous effects, such as convection currents in the air due to slight temperature differences. He wanted to make sure that he was not measuring magnetic forces by mistake. He needed many determinations to make sure that his results were reproducible and to determine their accuracy. With these experiments Cavendish determined G. He expressed his answer in terms of the mean density of the earth, which he found to be close to $5\frac{1}{2}$ times that of water, very close to Newton's estimate.



15–17 A simplified diagram of the apparatus used by Cavendish to verify the law of universal gravitation for small objects and to measure the gravitational constant G.

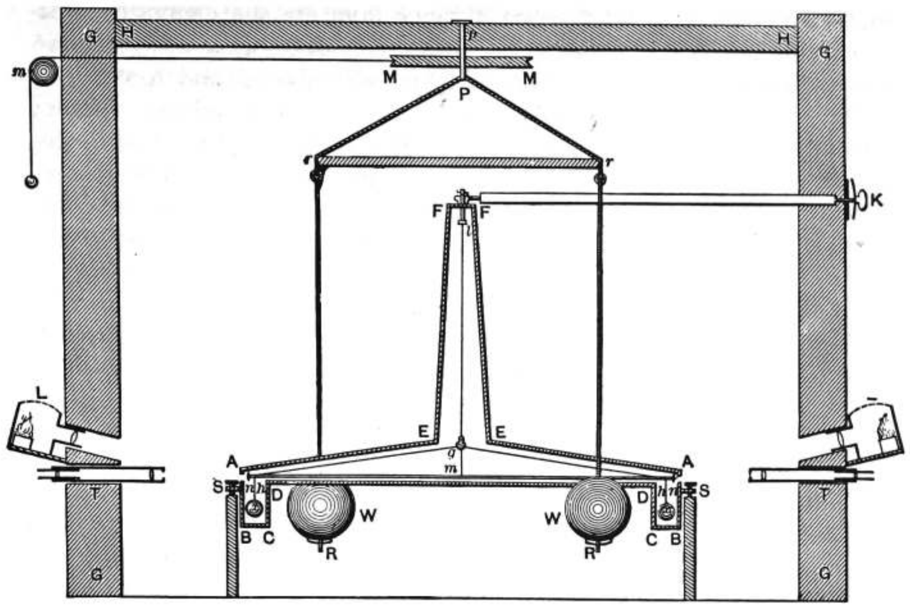
By using different substances for the various objects, we can modify the Cavendish experiment to show that the masses alone determine the gravitational attraction. By changing the relative positions of the small and large masses, we can check up on the inverse-square law at laboratory rather than planetary distances. Many modifications of the Cavendish experiments have been carried out, and so far Newton's law of universal gravitation is consistent with all of them.

15–11 A SMALL DISCREPANCY

Three centuries have passed since Newton's work on gravitation. During that time the law of gravitation has been tested through the most detailed calculations of the motion of the planets and their moons. In almost every case, the calculations have predicted orbits which are in agreement with actual observation. There is, however, one exception: an extremely small irregularity in the orbit of the planet Mercury which is not predicted by Newton's law of gravitation. Even though the discrepancy is minute, an improved theory is needed to explain it.

---

*The force at various twists of the suspension can be found dynamically by taking away the large masses and letting the rod carrying the small balls swing horizontally, twisting and untwisting the wire suspension. The motion of the rod then depends on the known masses of the small balls and on the forces exerted on them by the twisted suspension. The size of these forces can therefore be determined from the motion. Now consider the whole system at rest with the large masses in position. The net force on each ball must be zero; but the suspension is twisted because of the gravitational forces between the large masses and the balls on the rod. The net force, zero, is the resultant of the gravitational force and the force exerted by the twisted suspension. The gravitational force is therefore equal in magnitude and opposite in direction to the force exerted by the suspension. Since we now know the force exerted by the suspension, we also know the gravitational force.



15–18 The sketch of the Cavendish apparatus which appeared in Cavendish's original paper. The two large masses are labeled W and the small masses x. Notice that the whole device is mounted in a large case G with outside controls to move the weights and adjust the horizontal rod. Scales at A near the ends of the rod were illuminated by the lamps L and observed through the telescopes T.

Such a theory was provided by Albert Einstein in his general theory of relativity. At the heart of his theory lies the remarkable equivalence of gravitational and inertial mass. Einstein welded them into one entity. His theory is built upon the theory of Newton, just as Newton's theory was built on the work of Galileo and Kepler. Out of it come all the results of Newton's theory (but the calculations are more difficult). Actually, when we say it gives the same results, we mean that the differences between the predictions of Einstein's gravitational theory and Newton's mechanics are generally so small that they usually cannot be observed. Only under exceptional circumstances are the predicted differences observable. Mercury's orbit is one of these rare exceptions. Here Einstein's predictions of the orbit bring theory into accord with observation.

QUESTIONS

1. Three fireflies *X, Y*, and *Z* are on a moving bicycle at night. *X* is at the very center of one of the axles, which turns with the wheel. *Y* is on the rim of the wheel. *Z* is on the frame of the bike outside the circumference of the rim. Draw sketches and use a few words to describe the motions.
   (a) of *X* and *Y* as viewed by *Z*.
   (b) of *Y* and *Z* as viewed by *X*.
   (c) of *X* and *Z* as viewed by *Y*.
   (d) of all three as seen by an observer standing near the bicycle.

2.* (a) In Fig. 15 – 4, what is the ratio of the greatest distance to the shortest distance between the earth and Mars?
   (b) What about Fig. 15 – 7 (a)?

3. The fixed stars go around in about one day as seen by an observer on the earth. About how long do the following revolutions appear to take?
   (a) The sphere of the fixed stars as seen by someone on the moon. (From the earth, we always see the same side of the moon.)
   (b) The earth about the moon as seen from the moon.
   (c) The sun as seen from the moon. Does the sun appear to rotate as if it were one of the fixed stars? Does it appear to rotate faster or slower? Remember that the earth and moon go around the sun once every year.

4. Does Tycho Brahe's geocentric system account for the fact that there is a maximum angle between Venus and the sun as seen from the earth?

5. The earth's orbit around the sun is nearly circular, and the moon moves in a nearly circular orbit around the earth. The radius of the earth's orbit is 1.5×10^{11} meters, and that of the moon's orbit is 4×10^{8} meters.

   (a) How often is the moon between the earth and the sun?

   (b) How far has the moon moved around the sun in the interval between two successive times when the moon is between the earth and the sun?

   (c) Sketch the orbit of the earth going around the sun and on the same figure sketch the orbit of the moon going around the sun.

   (d) Does the moon show retrograde motion as seen by an observer at the sun? (Retrograde motion is motion like that shown in Fig. 15 – 2.)

6.* What area is swept out by Jupiter in making one fourth of a revolution around the sun? (See Table 2.)

7. How fast, in m^2/sec, is area swept out by
   (a) the radius from sun to earth?
   (b) the radius from earth to moon?

8.* Figure 15 – 10 shows the orbits of the planets about the sun. When does Pluto move (a) fastest? (b) slowest?

9.* Table 1 does not include Uranus, Neptune, and Pluto. What value do you predict for the ratio $\frac{R^3}{T^2}$ for each of these planets?

10.* If a small planet were discovered whose distance from the sun was eight times that of the earth, how many times longer would it take to circle the sun?

11. Between September 21 and March 21 there are three days fewer than between March 21 and September 21. These are the dates when day and night are of equal length, and between them the earth moves 180° around its orbit with respect to the sun. From this and Kepler's law of equal areas, explain how you can determine the part of the year during which the earth is closest to the sun.

12. Astronomers have observed that Halley's comet has a period of 75 years and that its smallest distance from the sun is 8.9×10^{10} meters, but its greatest distance from the sun cannot be measured because it cannot be seen. Use this information together with the second footnote in Section 15 – 5 to compute its greatest distance from the sun. (It was Newton who told Halley how to compute the orbit of a comet. Halley found and calculated the orbit and period of the comet that bears his name in the course of a general analysis he made of comets' orbits.)

13.* The radius of the moon's orbit is 60 times greater than the radius of the earth. How many times greater is the acceleration of a falling body on the earth than the acceleration of the moon toward the earth?

14.* (a) Calculate K_e from Table 2.

   (b) Using the value of K_s from Table 1 and your answer to (a), determine the ratio of the mass of the earth to the mass of the sun.

15.* At what height above the earth's surface will a rocket have half the force of gravitation on it that it would have at sea level? Express your answer in earth radii.

16. (a) At what height will a satellite moving in the plane of the equator stay over one place on the equator of the earth? One way to get the answer is to compare this satellite with the moon, which is at 59.5 earth radii from the center of the earth and takes 27 days to go around the earth.

   (b) How fast is the satellite accelerating toward the earth?

   (c) Using the inverse-square law and g at the surface of the earth, find the gravitational field strength at the height of the satellite. Compare with your answer to (b).

17. Find the weight of a 100-kg man on Jupiter.

18. A 70-kg boy stands 1 meter away from a 60-kg girl. Calculate the force of attraction (gravitational) between them.

19. A satellite circles the earth once every 98 minutes at a mean altitude of 500 km. Calculate the mass of the earth. The masses of planets are actually calculated from satellite motions, and one reason for establishing artificial earth satellites is to get a better value for the mass of the earth.

20. (a) If T is the period of a satellite circling around just above the surface of a planet whose average density is ρ, show that ρT^2 is a universal constant.

   (b) What is the value of the constant?

21. The earth is acted upon by the gravitational attraction of the sun. Why doesn't the earth fall into the sun? Be prepared to discuss your answer.

22. (a) What is the speed of the moon around the sun compared to that of the earth around the sun?

(b) If the earth could be removed suddenly without disturbing the motion of the moon, what would be the subsequent path of the moon?

(c) Calculate the ratio of the force of attraction exerted by the sun on the moon to the force exerted by the earth on the moon.

(d) Why does the sun not capture the moon, taking it away from the earth?

23. Assume the earth is perfectly round and has a radius of 6400 km.

(a) How much less does a man with a mass of 100 kg apparently weigh at the equator than at the poles because of the rotation of the earth?

(b) How fast would the earth have to spin in order that he would exert no force on a scale at the equator?

(c) How many times larger is the speed of rotation in (b) than the actual speed?

24. A 10,000-kilogram spaceship is drifting on a long mission toward the outer edge of the solar system. It has put out a small experimental satellite which revolves around it at a distance of 120 meters under their mutual gravitational attraction.

(a) What is the period of revolution of the satellite?

(b) What is the speed of the satellite?

25. An astronomer observes a planet with a small satellite revolving around it in a circular orbit of radius r with a period of time T.

(a) What is the mass of the planet?

(b) What is the acceleration of the satellite toward the planet?

(c) What is the gravitational force on the satellite?

(d) The astronomer measures the radius of the planet and finds it to be 1/10 of the radius of the orbit of the satellite. What is the gravitational field strength of the planet at its surface?

26. Two unequal masses m_1 and m_2, which are isolated in free space, attract each other with a gravitational force $F = G\frac{m_1 m_2}{R^2}$. What are the accelerations of m_1 and m_2? Is your answer in conflict with the justified expectation that an observer on m_1 will see that m_2 is coming toward him with the same acceleration as that with which an observer on m_2 is seeing m_1 coming toward m_2?

27. Astronomical observations indicate that the sun is describing a circular orbit around the center of our galaxy. The radius of the orbit is about 30,000 light-years ($= 2.7 \times 10^{20}$ m) and the period of one complete revolution is about 200 million years. In this motion the sun is acted on by the gravitational pull of the great quantity of stars lying inside its orbit.

(a) Calculate the total mass of these stars from the data given.

(b) How many stars of mass equal to the sun (2×10^{30} kg) does this represent?

16

Momentum and the Conservation of Momentum

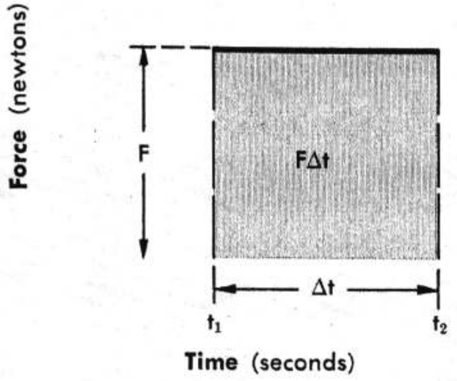
16–1 IMPULSE

Try to make a baseball and a 16-lb shot go at the same speed. As you know, it is a lot harder to get the shot going. If you apply a constant force $\vec{\mathbf{F}}$ for a time Δt, the change in velocity is given by $m\Delta\vec{\mathbf{v}} = \vec{\mathbf{F}}\Delta t$. So, in order to get the same $\Delta\vec{\mathbf{v}}$, the product $\vec{\mathbf{F}}t$ must be greater the greater the mass m you are trying to accelerate.

To start a 16-lb shot from rest and give it the same final velocity as a baseball (also started from rest), we must push either harder or longer. What counts is the product $\vec{\mathbf{F}}\Delta t$. This product $\vec{\mathbf{F}}\Delta t$ is the natural measure of how hard and how long we push to change a motion. It is called the *impulse* of the force.

We can exert a given impulse in many different ways: a big force for a short time, a smaller force for a longer time, or even a force that changes while it acts. In Fig. 16 – 1 we have plotted a constant force F against the time it acts. The plot is just a horizontal straight line of height F above the time axis and of length $\Delta t = t_2 - t_1$, equal to the time during which the force acts. The area of the rectangle under this line is $F\Delta t$, the size of the impulse in the time interval. (The direction of the impulse is the same as that of the force.) For any constant force acting for any time interval, we can always get the magnitude of the impulse as the area under the force-time curve for this time interval, and the direction is the direction of the force.

Now suppose the force changes from the constant value $\vec{\mathbf{F}}_1$ to the constant value $\vec{\mathbf{F}}_2$. For a time Δt_1 the force is constant and the impulse is $\vec{\mathbf{F}}_1\Delta t_1$. This impulse makes the change $m(\Delta\vec{\mathbf{v}})_1$ in the motion. In the next time interval, Δt_2, the force is again constant, this time $\vec{\mathbf{F}}_2$, and there is an impulse $\vec{\mathbf{F}}_2\Delta t_2$ which results in the change $m(\Delta\vec{\mathbf{v}})_2$ in the motion.



16–1 **A constant force F is plotted against time for the interval $t_2 - t_1 = \Delta t$. The area FΔt under the curve gives the impulse of this force in time Δt.**

That is,

$$\vec{F}_1\Delta t_1 = m(\Delta\vec{v})_1$$

and

$$\vec{F}_2\Delta t_2 = m(\Delta\vec{v})_2$$

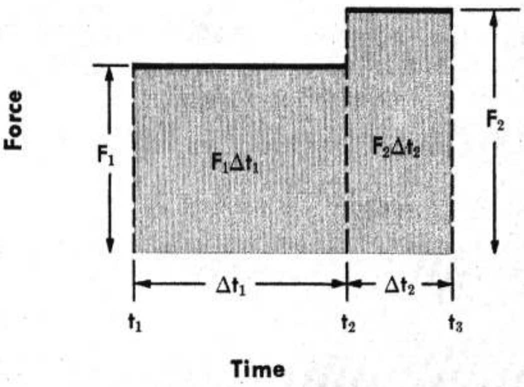
If we add these two statements vectorially we get

$$\begin{aligned}\vec{F}_1\Delta t_1 + \vec{F}_2\Delta t_2 &= m(\Delta\vec{v})_1 + m(\Delta\vec{v})_2\\ &= m[(\Delta\vec{v})_1 + (\Delta\vec{v})_2]\end{aligned}$$

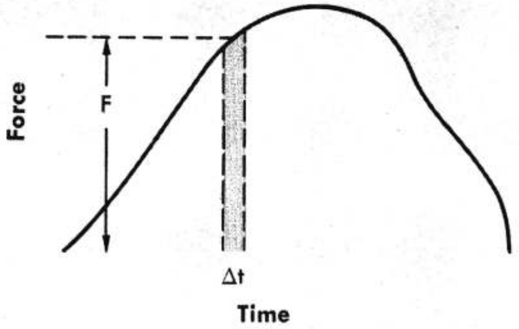
Thus, the total impulse during $\Delta t = \Delta t_1 + \Delta t_2$ just equals the mass times the total change in vector velocity that occurs during Δt.

When we deal with a continuously varying force we can still obtain the total impulse by adding up all the impulses in short time intervals. We can take each time interval so short that $\vec{F}$ is essentially constant during it. Then the total impulse, which is the sum of all the $\vec{F}\Delta t$'s, gives the total change $m\Delta\vec{v}$ in the motion.

When the force changes magnitude but always points in the same direction, we can add the impulses up by finding the area under the F–t curve. A simple case is shown in Fig. 16 – 2, and an example in which the force changes continuously in Fig. 16 – 3. If the force changes direction, however, this graphical method is not enough. We must add the little impulses $\vec{F}\Delta t$ in their correct directions, using vector addition to find the total impulse. If the first little impulse occurs at time t, when the moving object has velocity $\vec{v}$, and the final impulse occurs at time t' when the object has velocity $\vec{v}'$, then



16–2 **The force is F_1 in the time interval Δt_1, and changes to F_2 in time interval Δt_2. If the direction does not change, the total impulse in time $\Delta t_1 + \Delta t_2$ is $F_1\Delta t_1 + F_2\Delta t_2$, the shaded area under the curve.**



16–3 **When the force is continually changing, we can break up the area under the F-t curve into a large number of small areas that are nearly rectangles. (See shaded area.) Each such area FΔt is a little impulse. The total impulse is the sum of all these small impulses. Therefore, as long as the force points in the same direction, the area under the curve is the total impulse.**

$$\text{total impulse} = m(\vec{v}\,' - \vec{v}\,) = m\vec{v}\,' - m\vec{v}$$

(We shall often use the symbol ′, which is read "prime," to distinguish a quantity at a later time from that same quantity at an earlier time.)

Often a large force acts only for a very short time interval. Think what happens when you hit a tennis ball with a racket, or when two steel balls collide. Without the most elaborate equipment it is very hard to tell the size of the forces which act during the collision. But the total impulse of the force is easily found by observing the net change in $m\vec{\mathbf{v}}$. Forces which act for a short time and for which only the product $\vec{\mathbf{F}}\Delta t$ is known are called impulsive forces.

16–2 MOMENTUM

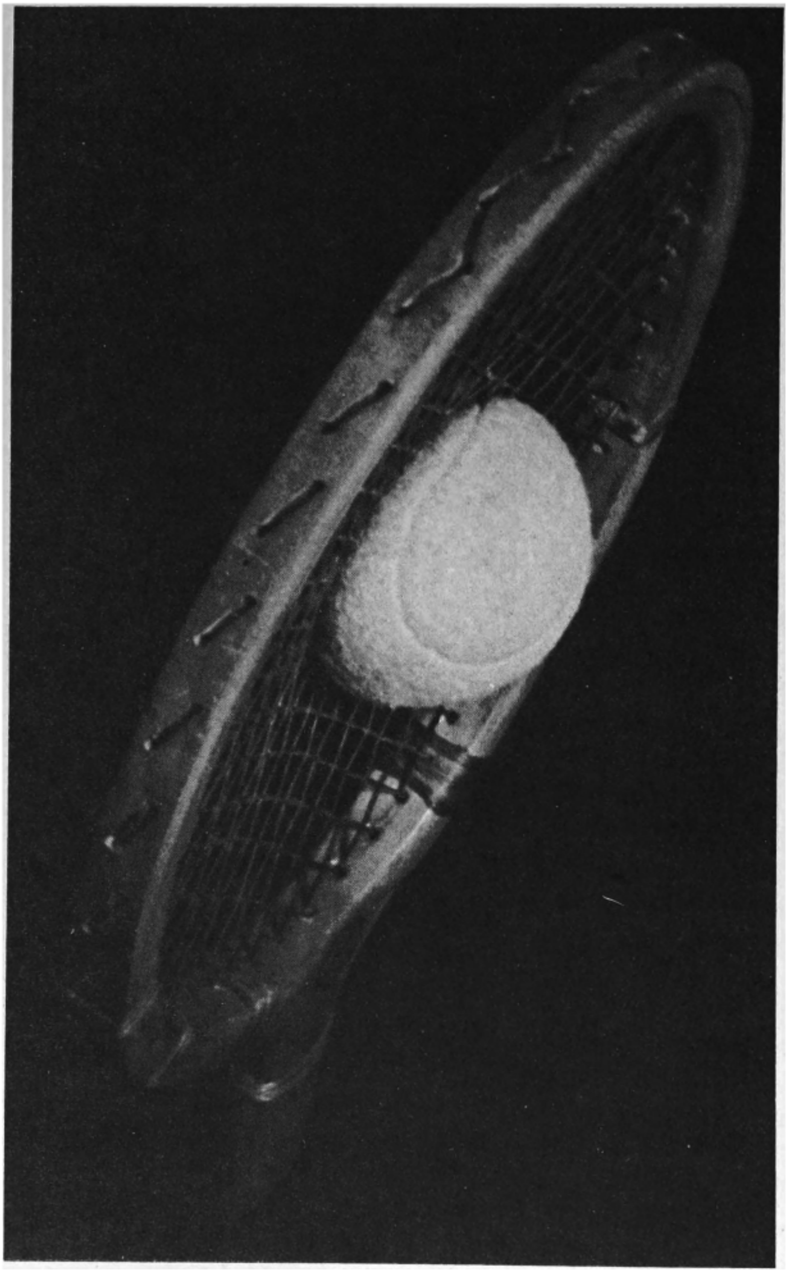
Suppose we apply the same impulse to two quite different objects, say a baseball and a 16-lb shot, both initially at rest. Since the initial value of the quantity $m\vec{\mathbf{v}}$ is zero in each case, and since equal impulses are applied, the final values $m\vec{\mathbf{v}}'$ will be equal for the baseball and the shot. Yet, because the mass of the shot is so much greater than the mass of the baseball, the velocity of the shot will be very much less than the velocity of the baseball. The product $m\vec{\mathbf{v}}$, then, is quite a different measure of the motion than simply $\vec{\mathbf{v}}$ alone, but it will prove to be of utmost importance in physics. We call it the *momentum* of the body, and measure it in kilogram-meters per second. Because we shall use momentum often, we introduce the symbol $\vec{\mathbf{p}}$ to represent the momentum of a body:

$$\vec{\mathbf{p}} = m\vec{\mathbf{v}}$$

Velocity and momentum, although related, tell us different things. The knowledge of the velocity only tells how fast (and in which direction) an object moves. It tells us nothing about the effort required to get it moving, or, for that matter, to stop it. The momentum, on the other hand, does not tell the speed of the object (although it tells us the direction in which it moves), but it determines the impulse necessary to get it going and the impulse necessary to stop it. Briefly, the velocity is a kinematical quantity—one which enters the geometrical description of where and when—whereas the momentum is a dynamical one, connected with the impulses and therefore with the causes of changes in the motion of masses.

Notice that the momentum does not depend on the way in which the body acquired its present motion. The equation $\vec{\mathbf{p}} = m\vec{\mathbf{v}}$ contains nothing but the mass and its motion now. The impulse which set it going may have been delivered in any of an infinite number of ways (as we know from the last section), or the mass may always have been moving with momentum $\vec{\mathbf{p}}$, for all we care. Similarly, the impulse required to bring the mass to rest may be applied in infinitely many ways or not at all; we still know from the present momentum how big it would be. It would cost a reverse impulse equal to $-m\vec{\mathbf{v}}$ to stop the body whether we use a big force and short time or a little force and long time.

Because of its connection with the impulse which occurs naturally in Newton's law $\vec{\mathbf{F}}\Delta t = m\Delta\vec{\mathbf{v}}$, we expect momentum to fit naturally into Newtonian dynamics. In fact, Newton



16–4 This high-speed photograph shows a familiar example of an impulsive force. Notice the large distortion of the racket and the ball. (Courtesy: Harold E. Edgerton.)

expressed his law of motion in terms of the momentum $m\vec{v}$ which he called the quantity of motion. We can easily express Newton's law in terms of the change in momentum instead of change in velocity:

$$\vec{F}\Delta t = m\Delta\vec{v} = m(\vec{v}' - \vec{v})$$

where $\vec{v}$ and $\vec{v}'$ are the velocities before and after the impulse $\vec{F}\Delta t$. But the right-hand side of the last equation can be written as

$$m(\vec{v}' - \vec{v}) = m\vec{v}' - m\vec{v} = \vec{p}' - \vec{p} = \Delta\vec{p}$$

the change in the momentum. Therefore

$$\vec{F}\Delta t = \Delta\vec{p}$$

or, in words, the impulse equals the change in the momentum. Then the average force which would produce the required impulse during Δt is just

$$\vec{F}_{av} = \frac{\Delta\vec{p}}{\Delta t}$$

If the actual force is changing, we can find its instantaneous value at any time provided we know how the momentum varies with time. We simply take Δt smaller and smaller until we can find a value for $\frac{\Delta\vec{p}}{\Delta t}$ which does not change much with further decreases in Δt. In other words, the instantaneous force is

$$\vec{F} = \lim_{\Delta t \to 0} \frac{\Delta\vec{p}}{\Delta t}$$

The force at any time equals the rate of change of the momentum. It was in this form rather than in terms of $\vec{F} = m\vec{a}$ that Newton originally formulated his law.

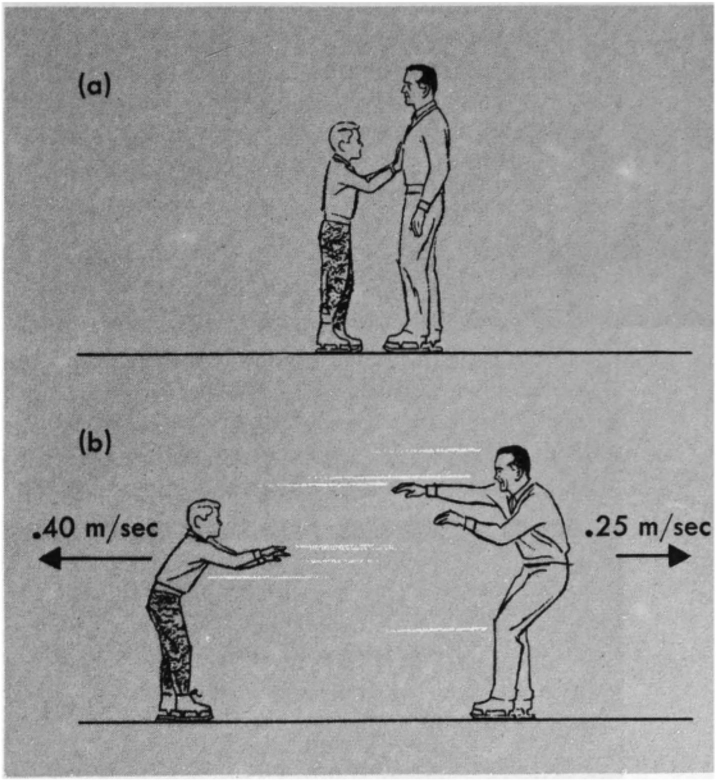
For the discussion of the motion of a single object, both forms of Newton's law are equally convenient. The great importance of momentum will become apparent in the next sections where we deal with the motions of two objects which exert forces on each other.

16–3 CHANGES IN MOMENTUM WHEN TWO BODIES INTERACT

A boy and a man are standing next to each other on a smooth sheet of ice. The boy shoves the man and they both start to move, sliding away in opposite directions with the boy moving away somewhat faster than the man. If we perform such experiments, we find that whenever two people are at rest and one pushes the other, they go off in opposite directions. We also find that their speeds are inversely proportional to their masses. For example, if a 50-kg boy shoves an 80-kg man hard enough so that the man goes off at .25 m/sec, we find that the boy goes off in the opposite direction at .40 m/sec (Fig. 16 – 5).

These experimental facts are most easily expressed in terms of the momentum of the man and the momentum of the boy. The momentum of the man is $mv = 80$ kg × 0.25 m/sec = 20 kg-m/sec, and the momentum of the boy in the opposite direction is $mv = 50$ kg × 0.40 m/sec = 20 kg-m/sec. After the shove their momenta are equal in magnitude and opposite in direction.

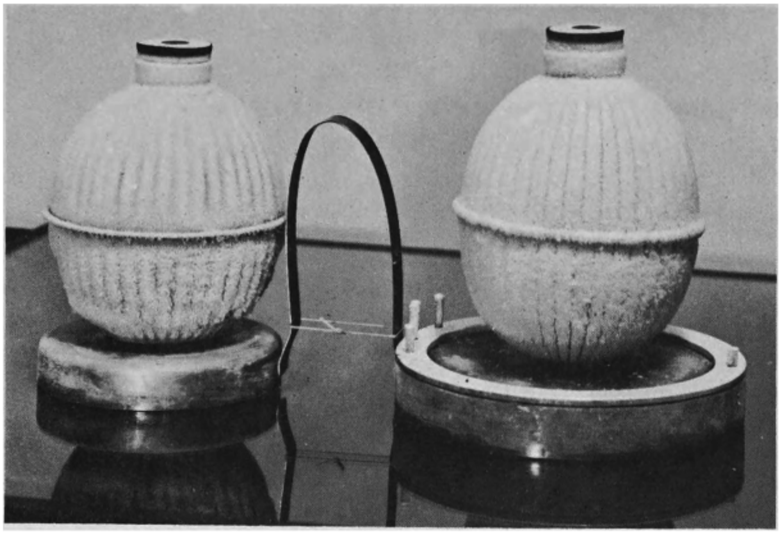
With Dry Ice pucks we can perform an accurate experiment of the kind that we have been describing. We screw one end of a spring to the edge of one puck; we bend the spring and tie



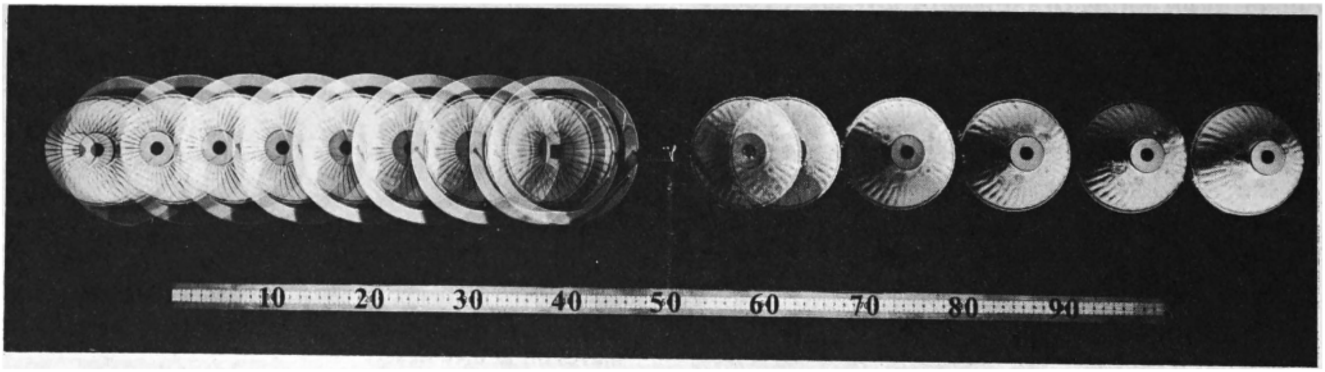
16–5 On a frozen pond a 50-kg boy pushes an 80-kg man hard enough to give him a speed of 0.25 m/sec. The boy moves off at 0.40 m/sec.

the ends together with a thread (Fig. 16 – 6). Then we put a second puck close to the first, as shown in the figure. When the thread is broken, the spring will snap out and shove the second puck away. When the two pucks are at rest on a level glass plate, we burn the thread and see what happens when the spring "explodes." The subsequent motion of the pucks is shown in Fig. 16 – 7.

In this experiment the mass of the large puck, including the spring mounted on it, was 3.9 kg, and the mass of the small puck was 2.0 kg. A glance at the picture shows that the two pucks went off in different directions. In each interval between flashes of the light the small puck moves farther than the large one. The small puck must have been given a higher velocity than the large one. By making measurements on the photograph you can tell that the small puck was moving at 0.48 m/sec after the shove, and the large puck was moving at 0.24 m/sec. The momentum of the small puck was $mv =$ 2.0 kg × 0.48 m/sec = 0.96 kg-m/sec to the right, while the momentum of the large puck was $mv =$ 3.9 kg × 0.24 m/sec = 0.94 kg-m/sec to the left. Within experimental accuracy these momenta are equal and opposite.



16–6 The apparatus used to show a simple kind of explosion. A compressed spring is held between two Dry Ice pucks. The spring is screwed to the larger puck and secured by a thread.



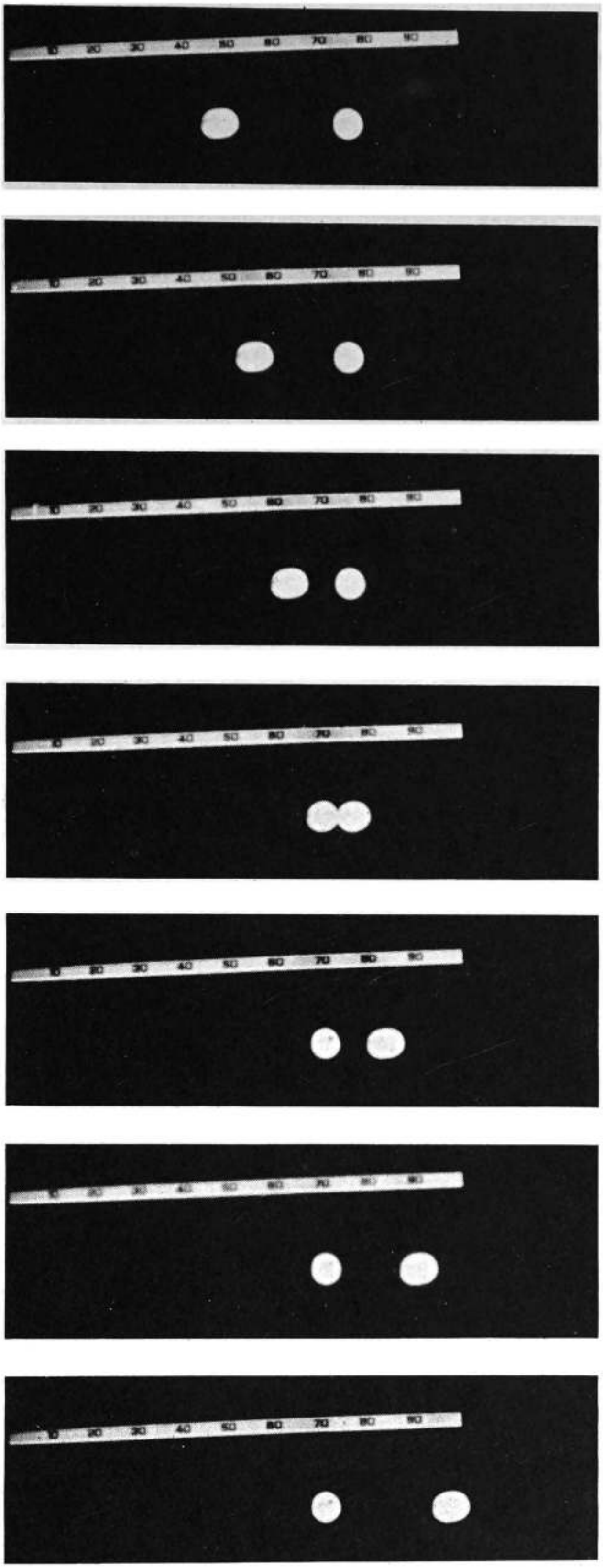
16–7 A multiple-flash picture of the motion of the two pucks after being pushed apart by the spring. The flash rate was 4 per second, and the scale is in centimeters.

By observing the motions of two bodies starting from rest and interacting with each other, we find time after time that the resulting changes in momentum are equal and opposite. It does not matter what produces the forces of interaction. They may arise from our muscles, from a distorted spring, or from a chemical explosion.

Take a chemical explosion as another example. When a man fires a rifle, the gases from the explosives exert violent forces inside the barrel. The bullet is pushed out in one direction and the rifle is pushed back in the other. We can show that the rifle's momentum and the bullet's momentum are equal and opposite: we suspend the rifle on long strings, take suitable flash photographs of the bullet, and time the very much slower motion of the recoiling rifle. In ordinary use the recoil velocity of the gun is rapidly decreased by an impulse from the man's shoulder, and a later measurement of its momentum does not show the equality and opposition so directly.

So far we have examined only cases in which both of the interacting bodies were originally at rest. What can we say about the changes in the momenta of two interacting bodies when one or both are originally in motion?

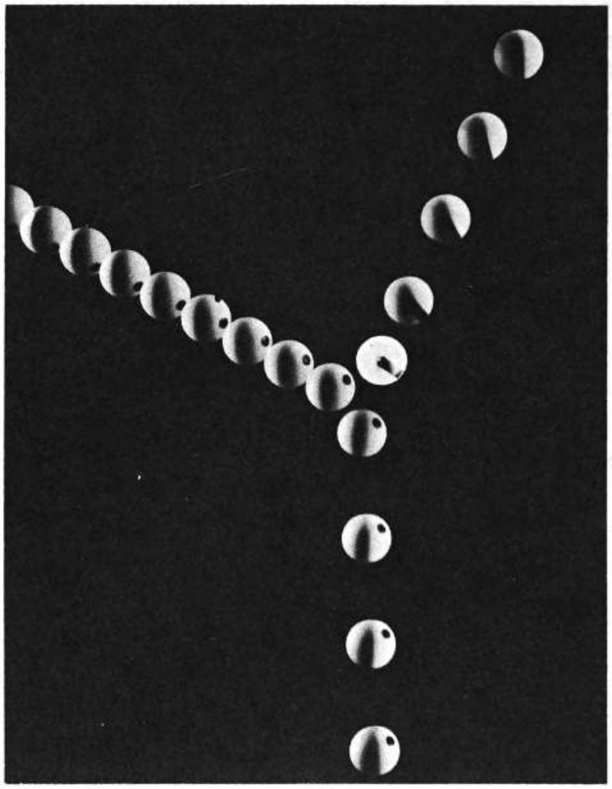
In Fig. 16 – 8 a moving billiard ball collides with a billiard ball at rest. The incident ball



16–8 Enlarged movie frames showing a head-on collision between two billiard balls of equal mass. The time interval between pictures is 1/48 second; the scale is in centimeters. Notice that all of the momentum of the incident ball is transferred to the ball initially at rest.

stops and the ball it hits goes off with the same velocity with which the incident ball came in. The two billiard balls have the same mass. Therefore, the momentum of the second ball after the collision is the same as that of the incident ball before collision. The incident ball has lost all its momentum, and the ball it struck has gained exactly the momentum which the incident ball lost. The changes in momentum are again equal and opposite.

The collision illustrated in Fig. 16 – 8 is a very special one: the incident billiard ball hit the ball at rest head on. Usually one billiard ball hits another somewhere on the side, and the two balls go off in different directions. Such a collision is illustrated in Fig. 16 – 9. In the picture the ball originally in motion came in at the bottom. After the collision, as we should expect, one ball goes off to the right; the other goes off to the left.



16–9 A multiple-flash photograph (flash rate 30 per second) of an offcenter collision between two balls, each of mass 173 grams. The dotted ball entered at the bottom of the picture and struck the striped ball at rest. (The camera was pointed straight down and the balls were moving horizontally.) To get the velocities and momenta diagramed in Fig. 16–10, we used the fact that the photograph is about 12% actual size. Could this approximation affect the conclusion that $\Delta\vec{p}_1 = -\Delta\vec{p}_2$?

The flash photograph shows the speeds and directions of the balls both before and after collision. The speeds are measured by the distances the balls travel between flashes. Using these measured speeds and the observed directions of motion, we find the velocity vectors, $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_1{}'$, and $\vec{\mathbf{v}}_2{}'$, representing the velocity of the first ball before and after the collision and that of the second ball after the collision. Then we obtain the momenta of the balls by multiplying the vector velocities by the masses: $\vec{\mathbf{p}} = m\vec{\mathbf{v}}$ for each ball. Because the masses of the balls here are equal, using $\vec{\mathbf{p}}$ instead of $\vec{\mathbf{v}}$ just changes the scale in this case. In Fig. 16 – 10 (a) we have plotted the momentum vectors $\vec{\mathbf{p}}_1$, $\vec{\mathbf{p}}_1{}'$ and $\vec{\mathbf{p}}_2{}'$, representing the momentum of the incident ball before the collision, and the momentum of each ball after the collision. (When the second ball was at rest before the collision, it had no momentum.)

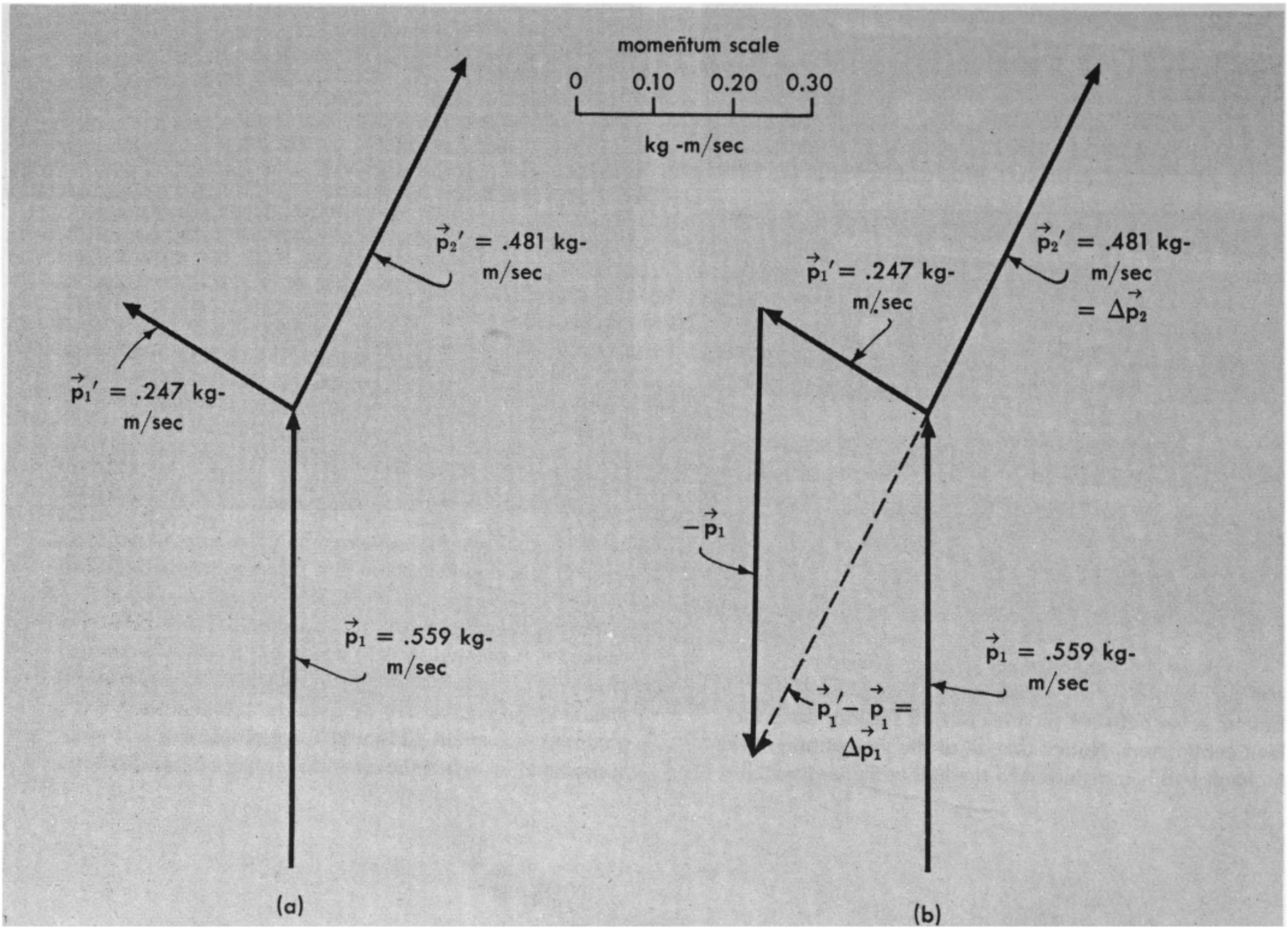
Now we shall find out whether the momenta of the two balls change by equal amounts in opposite directions in this experiment.

In Fig. 16 – 10 (b), $\Delta\vec{\mathbf{p}}_1$ is obtained graphically. We draw $-\vec{\mathbf{p}}_1$ from the head of $\vec{\mathbf{p}}_1{}'$. The vector $\Delta\vec{\mathbf{p}}_1$ is the resultant of $\vec{\mathbf{p}}_1{}'$ and $-\vec{\mathbf{p}}_1$. It extends from the tail of $\vec{\mathbf{p}}_1{}'$ to the head of $-\vec{\mathbf{p}}_1$, and we we see that it is equal and opposite to $\Delta\vec{\mathbf{p}}_2 = \vec{\mathbf{p}}_2{}'$, the change in the momentum of the struck ball. The change in momentum of the incident ball is compensated by an equal and opposite change in the momentum of the other ball, or

$$\Delta\vec{\mathbf{p}}_1 = -\Delta\vec{\mathbf{p}}_2$$

16–4 THE LAW OF CONSERVATION OF MOMENTUM

In the last section we have seen examples in which two bodies interact. The motion of each body changes, and the change in momentum of one body is equal and opposite to the change in momentum of the other. As we go further we

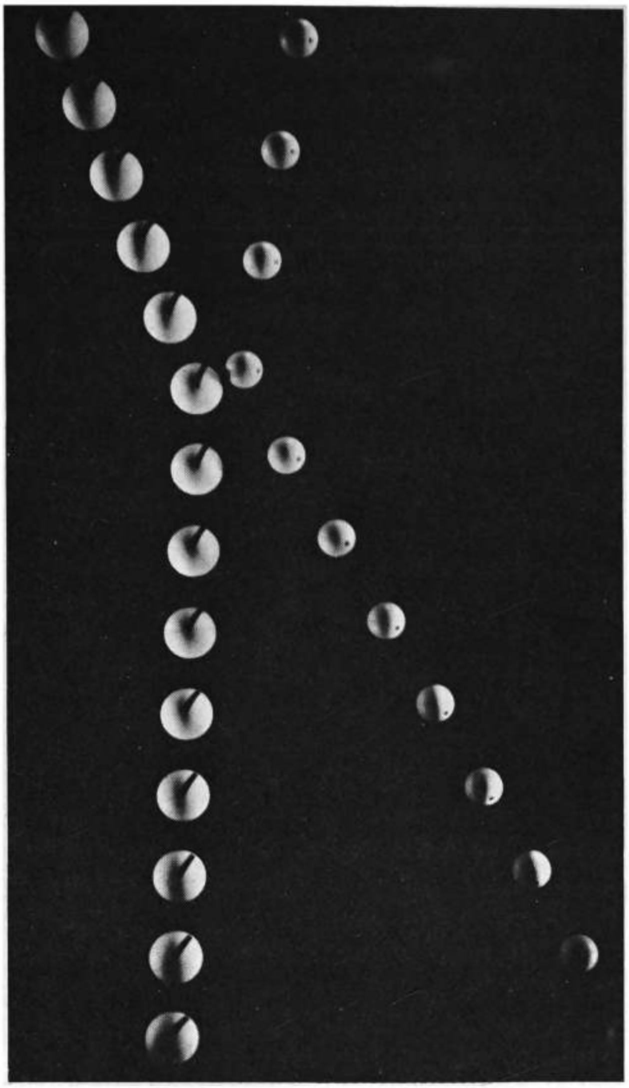

momentum scale
0
0.10
0.20
0.30
kg -m/sec
$\vec{p}_2{}' = .481$ kg-m/sec
$\vec{p}_1{}' = .247$ kg-m/sec
$\vec{p}_1 = .559$ kg-m/sec
(a)
$\vec{p}_1{}' = .247$ kg-m/sec
$\vec{p}_2{}' = .481$ kg-m/sec $= \Delta\vec{p}_2$
$-\vec{p}_1$
$\vec{p}_1{}' - \vec{p}_1 = \Delta\vec{p}_1$
$\vec{p}_1 = .559$ kg-m/sec
(b)

16–10 Below: (a) The vectors representing the momenta of the balls in Fig. 16–9. The value $\vec{p}_1$ is the original momentum of the dotted ball; $\vec{p}_1{}'$, its momentum after the collision; and $\vec{p}_2{}'$ the momentum acquired by the ball originally at rest. In (b) we obtain $\Delta\vec{p}_1 = \vec{p}_1{}' - \vec{p}_1$ graphically. We subtract $\vec{p}_1$ from $\vec{p}_1{}'$ by drawing $-\vec{p}_1$ from the head of $\vec{p}_1{}'$, then drawing the resultant (shown dashed in the figure). Note that $\Delta\vec{p}_1$ is closely equal and opposite to $\Delta\vec{p}_2 = \vec{p}_2{}'$.

shall come across more and more examples. No single example can yield a general proof that the momenta of two interacting bodies always change equally and oppositely. But our experience with the interactions of two bodies, in all circumstances, strongly suggests that equal and opposite changes of the momenta are the invariable rule in nature.

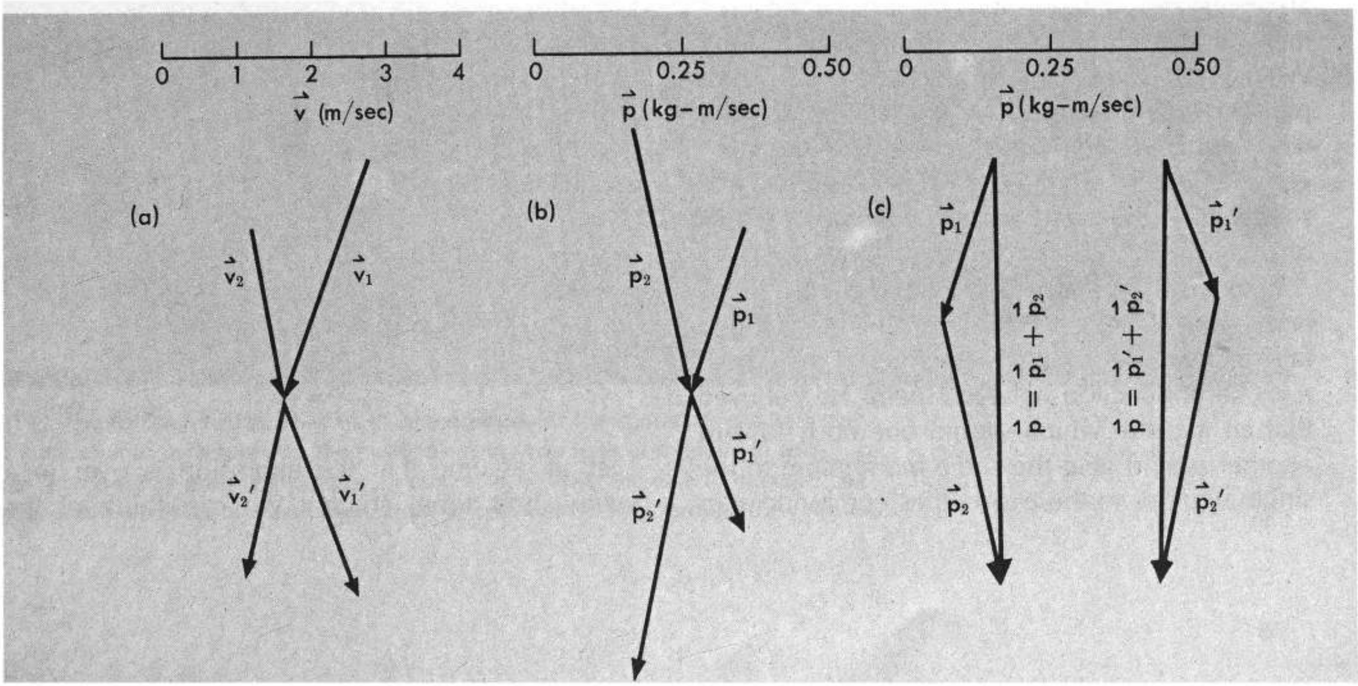
This rule of nature can be stated in a different form. We introduce the total momentum $\vec{P} = \vec{p}_1 + \vec{p}_2$ of the two bodies. Because any changes in the momenta $\vec{p}_1$ and $\vec{p}_2$ are just opposite, the total momentum $\vec{P}$ never changes. That is, $\Delta\vec{P} = 0$, or $\vec{P}$ is constant. We call the statement that the total momentum is constant the law of conservation of momentum.

Let us look at the conservation of momentum when two balls of different masses, both initially in motion, collide. Such a collision is illustrated in Fig. 16 – 11. By measuring the distances that the balls have moved between flashes, we determine their velocities. The velocity vectors are indicated in Fig. 16 – 12 (a). Multiplying by the masses, which were measured as 85.4 grams for the first ball and 201.1 grams for the second, we



16–11 A multiple-flash photograph of a collision between two balls – one of mass 201.1 g and the other 8.4 g. Again the flash rate was 30 per second and the setup was the same as that described in Fig. 16–9. Both balls were initially in motion and entered from the top of the picture.

16–12 Below: (a) The velocity vectors of the balls shown in Fig. 16–11. The subscript "$_1$" refers to the smaller ball, and "$_2$" to the larger. The primes indicate velocities and momenta after the collision. (b) The momentum vectors before and after the collision. (c) On the left we have added the momenta before the collision. On the right we have done the same for the momenta after. As you can see, $\vec{p}_1 + \vec{p}_2 = \vec{p}_1' + \vec{p}_2'$. In fact, when we made the original drawing, they were equal within about 1 part in 200.



obtain the momentum vectors for the individual balls before and after the collision [Fig. 16 – 12 (b)]. Then, on the left in Fig. 16 – 12 (c), we have added together the momentum vectors of both balls before the collision to obtain the total momentum; and on the right we have added together the momentum vectors of the two balls after the collision. We see that the two measured values of the total momentum are almost precisely the same. This experiment is consistent with the law of conservation of momentum.

Let us try one more kind of collision: one where the two colliding masses stick together and continue as a unit after the collision. Figure 16 – 13 shows a collision between a golf ball and a putty ball. The golf ball had a mass of 45.7 gm, and the putty ball, originally at rest, had a mass of 69.7 gm. The combined mass of the two balls stuck together after the collision was therefore 115.4 gm. This total mass was 2.53 times the mass of the incident ball. By measuring the ratio of the incident and final velocities in the flash photograph, you will find that the incident velocity was about 2.53 times the final velocity. Therefore, the final momentum was just equal to the initial momentum.

From now on we shall assume the conservation of momentum for two interacting bodies. We consider it to be a general law based on experience. Indeed, the kind of experience which we have summarized in our discussion here is the kind that originally led to the conservation law. Because the laws of collisions were of such obvious importance for the development of mechanics, in 1668 the Royal Society of London (then only eight years old) initiated an investigation into the problem. Three able men produced almost similar solutions: John Wallis, an English mathematician; Sir Christopher Wren, the English scientist and architect who designed St. Paul's; and Christian Huygens, a Dutch physicist. Wallis was first, and so he gets the credit for the general principle of conservation of momentum.

In a paper published in 1669, and later in his book "Mechanica" (1670), Wallis developed fairly clear ideas of impulse and momentum and the connection between them. He believed that an impulse which started one body toward another would give the same momentum to the single body or to the two bodies combined after collision. He thus arrived at the conservation of momentum for one kind of collision. He also argued that conservation applied to other collisions. All his arguments really involve assumptions about the forces or impulses between the colliding bodies. The French physicist Mariotte demonstrated these conclusions by a series of experiments on pendulums that could strike one another. Huygens made similar experiments, and other experiments in which hard spheres in close contact were struck by a similar sphere moving in a straight-line groove. In the "Principia" Newton described these and his own careful experiments. The law of conservation of momentum was established as one of the foundations of modern physics.

16–5 THE CENTER OF MASS

Think of a Fourth of July rocket which explodes while it is moving along through the air. After the explosion, if you saw all the fragments on one side of the path you would be sure that something was wrong. Your intuition leads you to expect that the fragments into which the rocket explodes should have an average motion along the original trajectory. In our investigation of such explosions we shall find that there is a special point that moves along in a constant way whether there is an explosion or not.

The fragments of an exploding skyrocket are a little too complicated for a first look at this effect. Let us go back to the simple explosion we have already discussed, the two Dry Ice pucks pushed apart by a compressed spring. By measuring the photograph (Fig. 16 – 7), we found that the two pucks, even though of quite different mass, had equal and opposite momenta after the explosion. If we call the puck on the left m_1 and the puck on the right m_2, we can write

$$m_1\vec{\mathbf{v}}_1 = -m_2\vec{\mathbf{v}}_2$$

Hence the ratio of the speeds is inversely proportional to the ratio of the masses:

$$\frac{v_1}{v_2} = \frac{m_2}{m_1}$$

Let us assume that the explosion is over in a very short time, so that we can think of the bodies as moving with constant speeds from the start. Then, after a time Δt, they will have moved distances of $x_1 = v_1\Delta t$ and $x_2 = v_2\Delta t$ from the point of the explosion. These two distances are directly proportional to the speeds, and hence are inversely proportional to the masses of the two bodies:

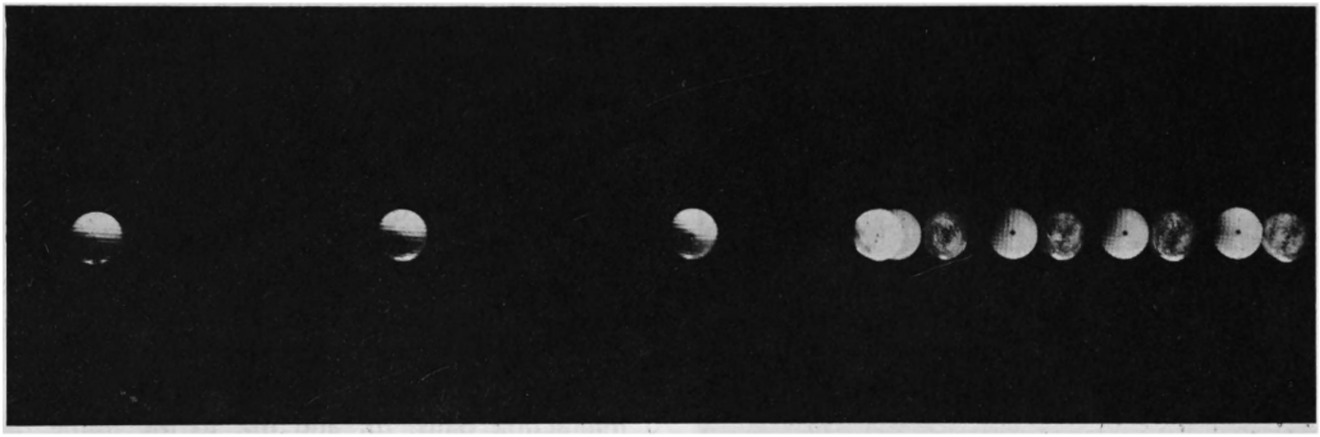
$$\frac{x_1}{x_2} = \frac{v_1}{v_2} = \frac{m_2}{m_1}$$

At all times during the motion, therefore, the starting point divides the distance between the two bodies in inverse proportion to their masses. We shall call the point which divides the distance between two bodies in the inverse ratio of their masses the *center of mass* of the two bodies. The starting point, in this case, is the center of mass.

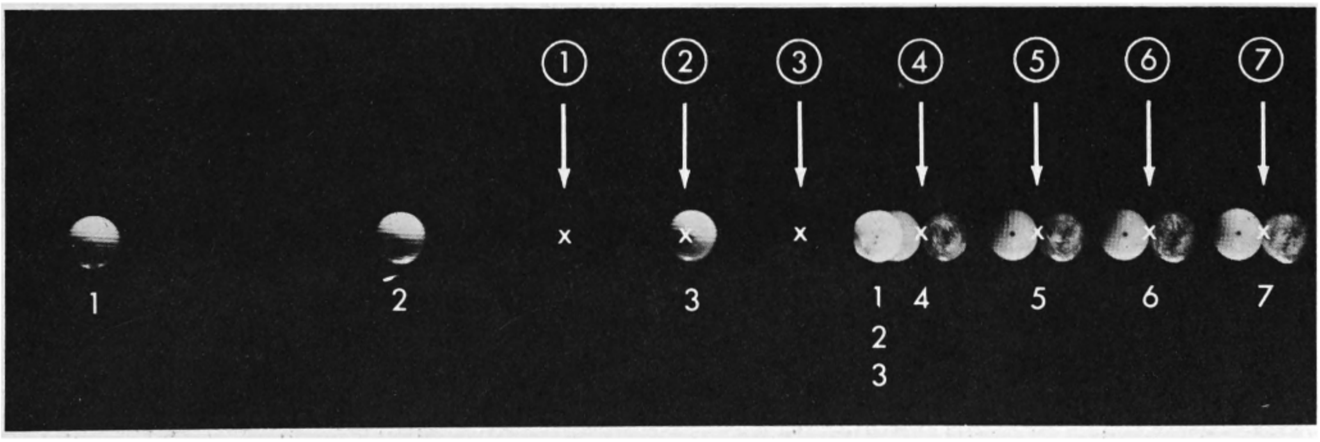
Let us investigate the behavior of the center of mass in some more complex interactions between bodies—the collisions we have discussed in this chapter. In Fig. 16 – 14 we again see the collision between a golf ball and a putty ball shown in Fig. 16 – 13 and analyzed in the last section. The masses were given as $m_1 = 45.7$ gm for the golf ball and $m_2 = 69.7$ gm for the putty ball. From the definition of the position of the center of mass we find

$$\frac{x_1}{x_2} = \frac{69.7}{45.7} \simeq \frac{3}{2}$$

That is, the center of mass divides the distance



16–13 A flash photograph of a collision between a golf ball and a putty ball. The golf ball moved in from the left and struck the putty ball. The two balls moved off to the right, stuck together. Measure the ratio of the incident and final speeds yourself, and check them against the calculation in the text.



16–14 The collision of Fig. 16–13 with the positions of the center of mass marked and numbered to correspond to the positions of the two balls at the time of each flash. The center of mass is located from $m_1x_1 = m_2x_2$, using the data of the text. You can see that the center of mass moves along at constant velocity through the collision.

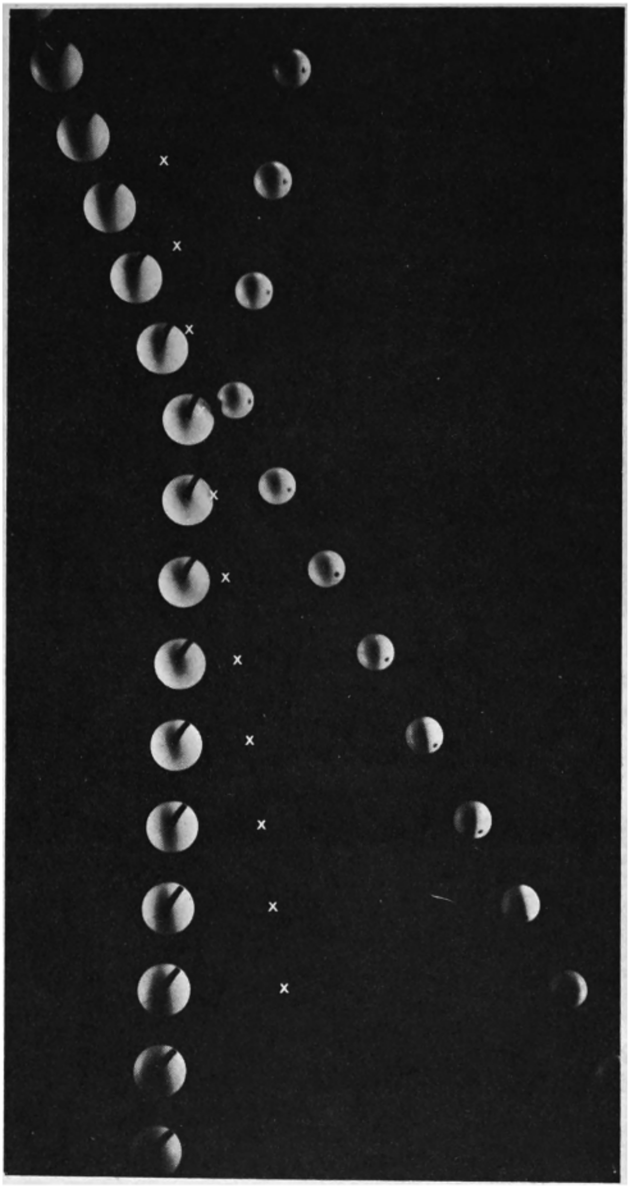
between the balls at any instant in the ratio of 3 to 2; it is to the right of the golf ball, just 3/5 of the separation distance at each flash. The positions of the center of mass at each flash are marked by arrows on the photograph and are numbered to correspond with the numbered positions of the balls. You can measure the velocity of the center of mass and see that it continues along in uniform motion, just as if no collision had occurred. In this case, where the two balls stick together, the velocity of the center of mass $\vec{v}_c$ is just the final velocity of the balls, $\vec{v}'$. By the law of conservation of momentum, the total mass $(m_1 + m_2)$ now has the same total momentum $\vec{p}$ as the incoming ball had before:

$$\vec{p} = (m_1 + m_2)\vec{v}' = m_1\vec{v}_1$$

But since $\vec{v} = \vec{v}_c$ we can also write

$$(m_1 + m_2)\vec{v}_c = m_1\vec{v}_1$$

Thus the total mass times the velocity of the center of mass gives the total momentum. We will see soon that the center of mass of two interacting bodies behaves that way even when the bodies do not stick together. The center of mass always moves as if all the mass were concentrated there.

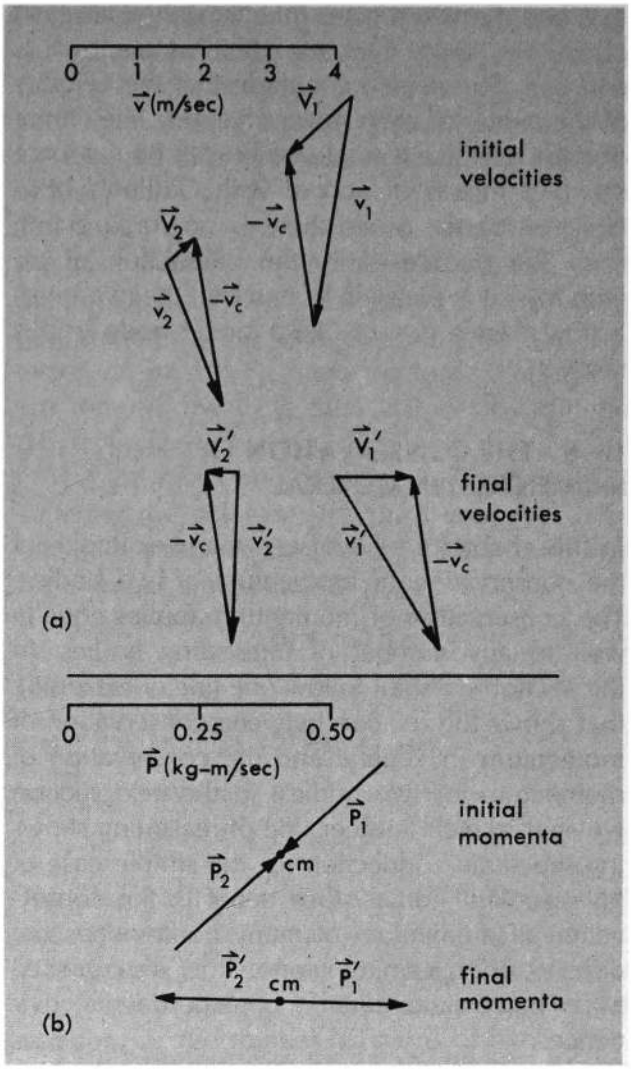


16–15 The collision of Fig. 16–11, showing the position of the center of mass at the time of each flash. The two balls had a mass ratio of about 7 to 3, so the distances between each ball and the center of mass are in the ratio of 3 to 7. Again the center of mass moves at constant velocity, quite undisturbed by the collision.

A more interesting collision is shown in Fig. 16 – 15. This is just the collision shown in Fig. 16 – 11 between two balls of masses 201.1 gm and 85.4 gm. Again we have marked and numbered the position of the center of mass for each flash. Even in this more complicated case, the center of mass continues right through the collision with constant velocity, along a perfectly straight line.

Clearly, the center of mass is a very special point. Let us move with that point, wherever it goes, and observe the motions of the two interacting balls, m_1 and m_2, in the frame of reference in which the center of mass is at rest. To get the velocities of the balls relative to the center of mass, we must subtract the velocity of the center of mass from the velocities of each of the balls. These velocities can be measured on Fig. 16 – 15 and are shown in Fig. 16 – 16 (a). As seen from the center of mass, the velocities $\vec{V}_1$ and $\vec{V}_2$ of the balls before collision are exactly opposite in direction and their magnitudes are inversely proportional to the masses. The same is true of the velocities $\vec{V}_1'$ and $\vec{V}_2'$ after the collision. Multiplying each of the velocities by the corresponding mass, we get the momentum diagram shown in Fig. 16 – 16 (b). It is clear that the momenta of m_1 and m_2 are equal and opposite: the total momentum of the balls, in the center-of-mass frame of reference, is always zero.

Now let us get down off the center of mass and return to earth. The velocity $\vec{v}_1$ of m_1, seen by an observer on earth (for example, the cam-


0
1
2
3
4
$\vec{v}$ (m/sec)
$\vec{V}_1$
initial velocities
$\vec{v}_1$
$-\vec{v}_c$
$\vec{V}_2$
$\vec{v}_2$
$-\vec{v}_c$
$\vec{V}_2'$
$\vec{V}_1'$
final velocities
$-\vec{v}_c$
$\vec{v}_2'$
$\vec{v}_1'$
$-\vec{v}_c$
(a)
0
0.25
0.50
$\vec{P}$ (kg-m/sec)
$\vec{P}_1$
initial momenta
$\vec{P}_2$
cm
$\vec{P}_2'$
cm
$\vec{P}_1'$
final momenta
(b)

16–16 (a) The velocity vectors of the balls and of the center of mass, measured from Fig. 16–15 and plotted. The subscript "$_1$" refers to the small ball, and "$_2$" refers to the larger. The primes indicate velocities after the collision. The vectors marked with capital letters are the velocities of the balls from the frame of reference of the center of mass. They are found by subtracting the velocity of the center of mass from the velocities of the balls. They are opposite in direction and the ratio of their magnitudes is just the inverse of the ratio of the masses of the corresponding balls. (b) The products mV, the momenta relative to the center of mass, are shown. They are equal in magnitude and opposite in direction, both before the collision and after the collision. The total momentum relative to the center of mass is always zero.

era used to take the photograph), is made of the velocity $\vec{v}_c$ of the center of mass plus the velocity $\vec{V}_1$ of m_1 with respect to the center of mass:

$$\vec{v}_1 = \vec{v}_c + \vec{V}_1$$

Similarly

$$\vec{v}_2 = \vec{v}_c + \vec{V}_2$$

Therefore the total momentum of m_1 and m_2 with respect to the earth is

$$\vec{p} = m_1\vec{v}_1 + m_2\vec{v}_2 = m_1(\vec{v}_c + \vec{V}_1) + m_2(\vec{v}_c + \vec{V}_2)$$

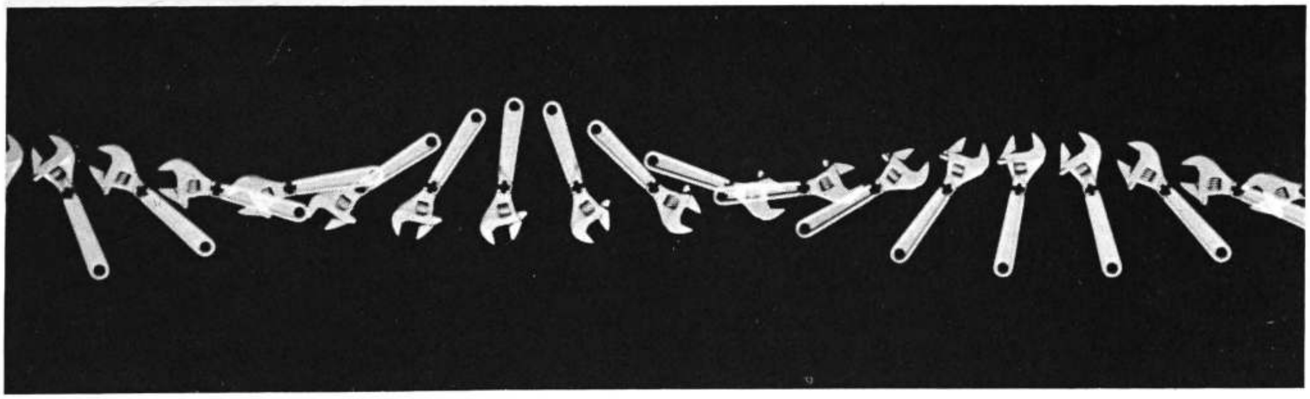
or $$\vec{p} = (m_1 + m_2)\vec{v}_c + m_1\vec{V}_1 + m_2\vec{V}_2$$

But by riding with the center of mass, we found that $m_1\vec{V}_1 + m_2\vec{V}_2$ is always zero. So finally we have

$$\vec{p} = (m_1 + m_2)\vec{v}_c$$

Once again, we find that the total momentum is the same as if the total mass moved with the center of mass.

The parts of a body may move with respect to each other. They may move around the center of mass or they may move in and out, toward and away from the center of mass. No matter what these internal motions, the center of mass of the body moves as though all the mass were there (Fig. 16 – 17).



16–17 A multiple-flash photograph (1/30 second between flashes) of a moving wrench. The black cross marks its center of mass.

When the two masses interact only with each other, we know that the total momentum is constant. So we also conclude that the velocity of the center of mass never changes. The center of mass acts like a single body with no net force on it. It moves in accord with Galileo's principle of inertia. Since there is no force acting from the outside—only the interaction of m_1 with m_2—it is pleasing to find that the two interacting masses do act like a single body in this respect.

16–6 THE CONSERVATION OF MOMENTUM IN GENERAL

In this chapter we have seen many examples of the conservation of momentum of two bodies. The conservation of momentum applies equally well to any number of interacting bodies. In this section we shall follow one line of reasoning that shows the relation between conservation of momentum in general and the conservation of momentum for two bodies. In the next section we shall sketch another line of reasoning showing the same connection for the simple case of "Newtonian" forces. Our belief in the conservation of momentum of many bodies does not depend on reasoning alone. It is also backed by a huge accumulation of experimental evidence.

In Fig. 16 – 18 we see a diagram of two bodies. One is a single ball and the other is made of two balls which are tied together by a very light spring. Suppose that the second body is at rest, and the first one hits it, moving with momentum $\vec{\mathbf{p}}_1$. After the collision, ball No. 1 travels away with a momentum $\vec{\mathbf{p}}_1'$ while the compound body moves off with momentum $\vec{\mathbf{p}}_2'$.

We know from the experiments we have been discussing that the momentum of two interacting bodies is conserved. Here, before the collision, the total momentum $\vec{\mathbf{P}}$ is just the momentum $\vec{\mathbf{p}}_1$ of the first body. After the collision the total momentum $\vec{\mathbf{P}}$ is $\vec{\mathbf{p}}_1' + \vec{\mathbf{p}}_2'$. Consequently

$$\vec{\mathbf{p}}_1 = \vec{\mathbf{p}}_1' + \vec{\mathbf{p}}_2'$$

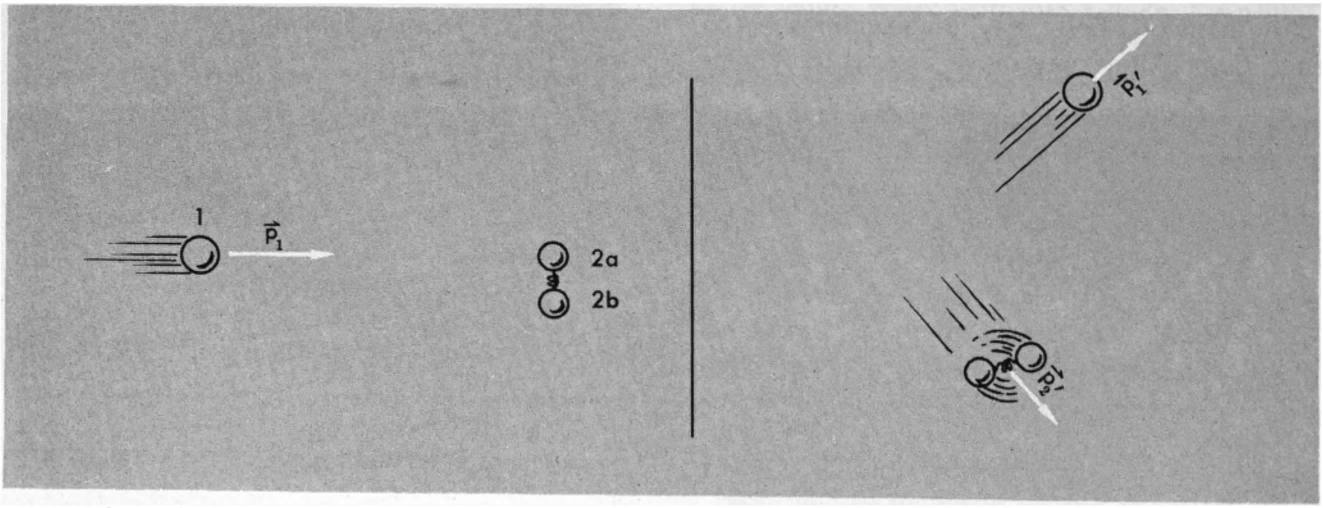
All we have said so far comes from considering the two bodies 1 and 2. Now let us take a closer look at body 2. It is made of the balls 2*a* and 2*b*. We know from the last section that its center of mass momentum $\vec{\mathbf{p}}_2'$ is the sum of the momenta of the two balls of which it is made:

$$\vec{\mathbf{p}}_2' = \vec{\mathbf{p}}_{2a}' + \vec{\mathbf{p}}_{2b}'$$

If we put this expression for $\vec{\mathbf{p}}_2$ into the preceding equation we get

$$\vec{\mathbf{p}}_1 = \vec{\mathbf{p}}_1' + \vec{\mathbf{p}}_{2a}' + \vec{\mathbf{p}}_{2b}'$$

This reasoning shows that the total momentum is conserved for three bodies (1, 2*a*, and 2*b*) just as well as two—at least when two of the three bodies are tied together. But the fact that balls 2*a* and 2*b* were connected by a spring really does not matter; the spring never entered the reasoning at all. The center-of-mass momentum $\vec{\mathbf{p}}_2'$ of body 2 is constant and equal to $\vec{\mathbf{p}}_{2a}' + \vec{\mathbf{p}}_{2b}'$ no matter if there is a spring holding the balls together or not.



16–18 The single ball, body No. 1, collides with body No. 2, which is made of two parts, 2a and 2b, connected by a light spring.

The experiment we just analyzed is no different from some of the other experiments we have considered. Figure 16 – 9 is a picture of a moving body colliding with one at rest. If you take a close look at that picture you can see that after the collision the ball that was hit rotates in addition to moving along as a whole. The momentum that we measured for the struck ball in Fig. 16 – 9 was really the momentum of the center of mass of that ball. Just as we can think of the collision that we discussed in the last paragraph as a collision between three bodies rather than two, we could have called the collision in Section 16 – 3 a collision of three bodies by mentally dividing the struck ball into two halves; for instance, the half with the black mark on it and the one without. Because of the forces holding the two halves together, these two halves are two bodies interacting with each other. After all, what we call a body is our choice. We decide whether to call that collection of arms and legs a chair or to split it into eight bodies called arms, back, seat, and legs.

By looking at the balls 1, 2*a*, and 2*b* in Fig. 16 – 18 as a system of three bodies, we showed that when three bodies interact the total momentum stays the same. The conservation of momentum applies just as well to three bodies as it did to two. We can now see that the conservation of momentum does not depend on the number of bodies at all. If we have any system of pieces that interact only among themselves, any isolated system, we can divide it into any number of bodies we choose. The same kind of reasoning that we just went through will show us that the total momentum does not change and does not depend on what we decide to call a body. The conservation of momentum therefore applies to any isolated system. It appears to be a perfectly general law. All experience, with interactions of any sort and any number of bodies, bears out this powerful conclusion.

16–7 FORCES OF INTERACTION

We can connect the conservation of momentum for two interacting bodies with the general law of conservation of momentum in another way. This line of reasoning involves the forces. We show that the forces in an interaction of two bodies are equal and opposite. Then we consider a number of bodies interacting and assume that the forces between every pair of bodies are still equal and opposite. With this assumption we find the total momentum of all the bodies is conserved.

In every interaction between two objects that we have studied, momentum is conserved; what one mass loses the other gains: $\Delta \vec{\mathbf{p}}_1 = -\Delta \vec{\mathbf{p}}_2$. Furthermore if we indicate the force that m_1 exerts on m_2 by $\vec{\mathbf{F}}_{1-2}$ and the force that m_2 exerts on m_1 by $\vec{\mathbf{F}}_{2-1}$, then $\Delta \vec{\mathbf{p}}_1 = \vec{\mathbf{F}}_{2-1}\Delta t$ and $\Delta \vec{\mathbf{p}}_2 = \vec{\mathbf{F}}_{1-2}\Delta t$.

The Δt must be the same for both masses suffering the collision. If I push you, you automatically push me back and I cannot push you for a longer time than you push me. Thus

$$\Delta \vec{\mathbf{p}}_1 = -\Delta \vec{\mathbf{p}}_2$$

becomes

$$\vec{\mathbf{F}}_{2-1}\Delta t = -\vec{\mathbf{F}}_{1-2}\Delta t$$

and that now leads to

$$\vec{\mathbf{F}}_{2-1} = -\vec{\mathbf{F}}_{1-2}$$

The idea that forces of interaction must be equal and opposite was stated by Newton (after his experiments on momentum conservation): "To every action there is always opposed an equal reaction, or the mutual reactions of two bodies upon each other are always equal and directed to contrary directions." This statement is often called his third law of motion, although you can see that it is a law about the forces in interactions and that the motions are inferred from it by applying Newton's law of motion to each body.

Even before we studied momentum, we studied one force of interaction with some care. In the last chapter we saw that the gravitational attraction between two bodies applies equal forces to each of the bodies, and the forces are directed oppositely along the line between the centers. The pull of the sun on the earth is equal and opposite to the pull of the earth on the sun. The pull of the earth on the moon is opposite to the force of the moon on the earth, and so on. Later in this volume we shall often study what goes on when the forces are of this particularly simple kind. We shall call them Newtonian forces.

Now let us trust Newton's third law to relate the forces between two bodies even when many bodies interact. Then we can use it to give an alternative demonstration of "universal" momentum conservation. We have a number of masses moving about in any way in an isolated system, exerting various forces one on another. Since all the forces occur in pairs with $\vec{\mathbf{F}}_{2-1} = -\vec{\mathbf{F}}_{1-2}$, their effects on the momenta of the bodies also occur in equal and opposite pairs. Therefore, while one body gains some momentum, another loses an equal and opposite amount. And in the whole isolated system there can be no net change of momentum.

If Newton's third law is true, momentum conservation follows. But it is possible that momentum may be conserved for many bodies even if some of the forces are not Newtonian. In fact, as we shall see in the latter part of this book, some forces are not Newtonian.

Here is one example in which the forces are not obviously Newtonian. Every once in a while a tremendous bright light flares up somewhere out in space—a supernova. The light gets dimmer exponentially, falling off by the factor $\frac{1}{2}$ every 55 nights (with the same half-life as the new radioactive element Californium). Compared to the millions of years light takes to reach us from a supernova, the light takes only a short period to die down. The supernova has burned out long before the light gets here.

When the light arrives here from a supernova, it gives us a slight shove. But nobody would think that we give the dead supernova a countershove at the same moment. If we think of forces of interaction between us and the supernova at the same moment, there is no reason to expect them to be equal and opposite.

The conservation of momentum is all right, however. What we must do is to include the light as one of the bodies in our description of the universe. When the light was emitted, the star gave it a shove and was given a shove backward. Some of the light took off toward us carrying momentum; and when it arrived here millions of years later, the light shoved on us and we shoved on it, so that as the light was absorbed here its momentum was transferred to us.

This description may not seem so farfetched if we return to the people sliding around on the ice. This time we imagine them playing ball. The man throws the ball to the boy. When he throws it, he gives it momentum in one direction and he receives momentum in the opposite direction. Later on, when the boy catches the ball, the momentum is transferred from the ball to him. (Of course the mass of the ball is included with him after he catches it.) The changes in the momenta of the man and of the boy with the ball are equal and opposite. And the interaction between the man and the boy occurs with a time delay. The time delay is considerably shorter than that which occurs between the emission of light by a supernova and its absorption at the surface of the earth; nevertheless, it is an appreciable and measurable time delay. If we forget about the ball, the interaction between the man and the boy does not appear Newtonian; and momentum seems to be lost for a while, when the ball is in the air. But everything makes sense when we include the ball in the system; and in just the same way everything makes sense in the case of the earth and the supernova when we include the light.

Looking at light as a ball thrown across from man to boy helps to make the time delay in the interaction less mysterious; and it gives a sense of assurance that momentum is still conserved in the overall interaction. The ball-throwing scheme is a "model" for the transfer of light—a model in which we can use Newtonian interactions. In fact, light is not a ball—balls do not appear when thrown and disappear when stopped, but light does. So we should not take the model too literally. When we study interactions with light, magnetic fields, etc., later on, we could continue to devise models to keep our description Newtonian in appearance; but the models are apt to be too complicated and to look artificial. We usually drop them and just hold on to the overall conservation of momentum.

By examining the emission and absorption of light (as we shall do to some extent later on), we shall find that the idea that momentum is carried by light makes a consistent picture. We can, and indeed we must, extend the law of conservation of momentum beyond those interactions in which the momentum is solely carried by material objects. We must include the momentum traveling around in radiation. The inequality in the forces between earth and supernova occurs while momentum is moving in radiation. There are many other similar examples. The same sort of thing happens when radiation is emitted by one atom and absorbed by another. When the momentum in radiation is included, the conservation of momentum spans our universe.

Rockets

We can use the law of conservation of momentum to understand the motion of a rocket, and from this we can predict some of the requirements that have to be satisfied to put an earth satellite into orbit.

Consider a rocket moving far out in space. We will assume that there are no forces acting on the rocket and that the velocity of the earth does not change appreciably during the time the rocket accelerates (a good approximation for the earth for a short period of time). In this case, the velocity $\vec{v}_0$ of the rocket with respect to the earth is constant.

The rocket engines are now turned on, and mass is thrown out in the form of a continuous stream of gas at a high constant velocity $\vec{v}_e$ with respect to the rocket. If $\vec{v}_e$ is opposite in direction to $\vec{v}_0$, the gas will move slower relative to the earth and have less momentum than when it traveled at velocity $\vec{v}_0$ before it was expelled. At the same time, the rocket will gain momentum in the direction in which it was originally moving.

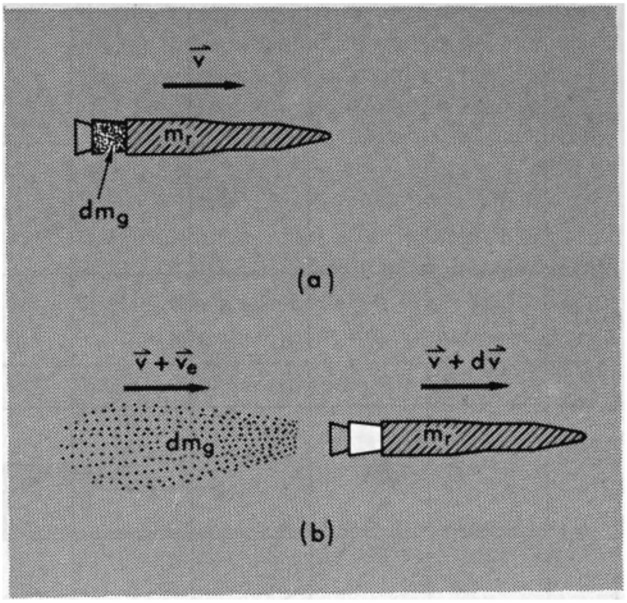
Let us consider the momentum of the rocket and the exhaust gas at the beginning and the end of a small time interval dt. At the start of the interval, the total momentum of the system is the sum of the momentum of the rocket of mass m_r and the momentum of a small mass of exhaust gas, dm_g, about to be ejected. Both m_r and dm_g are initially moving at velocity $\vec{v}$ with respect to the earth [Fig. 16 – 19 (a)], so the initial momentum is $m_r\vec{v} + dm_g\vec{v}$. At the end of the interval dt (Fig. 16 – 19b) the rocket is moving at a velocity $\vec{v} + d\vec{v}$, and the exhaust gas has slowed as a result of being ejected out the rear of the rocket to a velocity $\vec{v} + \vec{v}_e$. ($\vec{v}_e$ is negative relative to the earth.) The final momentum is therefore $m_r\,(\vec{v} + d\vec{v}) + d\mathbf{m}_g\,(\vec{v} + \vec{v}_e)$. Since momentum is conserved, the initial and final momentum are equal, so

$$m_r\vec{v} + dm_g\vec{v} = m_r\,(\vec{v} + d\vec{v}) + dm_g\,(\vec{v} + \vec{v}_e)$$

or

$$d\vec{v} = \vec{v}_e\,\frac{-dm_g}{m_r}$$

Notice that dm_g is the change in mass of the *gas*, and m_r is the mass of the *rocket*, so dm_g is not the differential of m_r but rather the differential of the variable $m_0 - m_r$, the total mass of exhaust gas ejected after a time t. (m_0 is the total initial mass of the exhaust gas and rocket before the engines are turned on.) Therefore, before integrating the above expression we must make a change of variable. The change in mass of the rocket is



16–19 Initially both the exhaust gas dm_g about to be ejected in a short time interval dt and the rocket mass m_r move at a velocity $\vec{v}$ relative to the earth. At a time dt later the rocket is moving at a greater velocity $\vec{v} + d\vec{v}$ and the exhaust gas is moving at a slower velocity $\vec{v} + \vec{v}_e$. (The exhaust gas is shot out of the rear of the rocket with a velocity $\vec{v}_e$ with respect to the rocket, so $\vec{v}_e$ is negative relative to the earth.)

negative but equal in magnitude to the change in mass of the exhaust gas moving at velocity $v + v_e$ (which steadily increases and is positive) so $dm_r = -dm_g$ and

$$d\vec{v} = \vec{v}_e \frac{dm_r}{m_r}$$

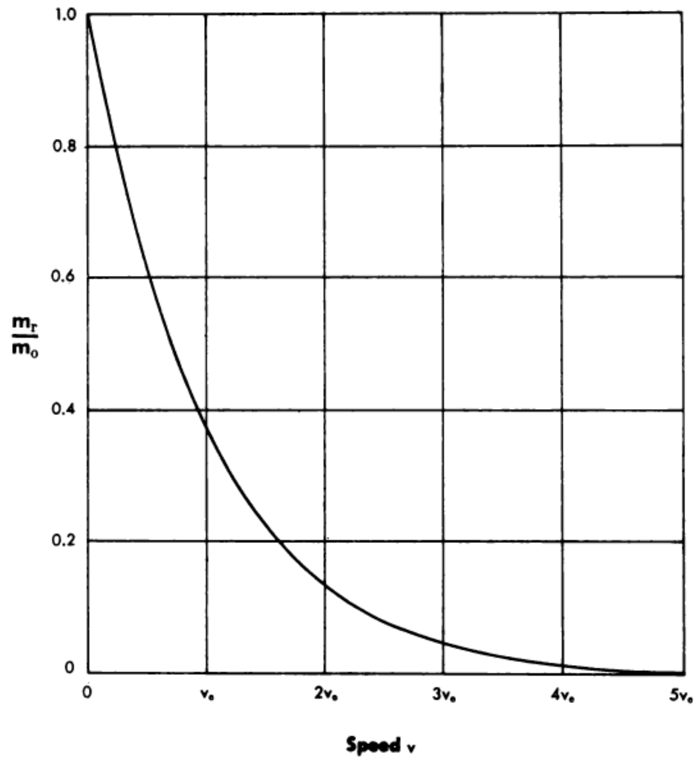
Taking components of the vectors and integrating, we have

$$\int_{v_0}^{v} dv = v_e \int_{m_0}^{m} \frac{dm_r}{m_r}$$

$$v - v_0 = v_e (\ln m_r - \ln m_0) = v_e \ln \left(\frac{m_r}{m_0}\right)$$

the change in velocity of the rocket as a function of the change in mass of the rocket after the motors were turned on. Expressed in exponential form, this becomes

$$\frac{m_r}{m_0} = e^{\frac{v - v_0}{v_e}}$$



16–20 A rocket throws out mass to gain speed. The fraction of the original mass remaining is plotted as a function of the speed v attained.

A graph of this relationship is shown in Fig. 16 – 20*. Note that v_e is negative.

We now can estimate how much of the mass of a rocket we must throw away to get an earth satellite into orbit. We already know that we need a satellite velocity of about 8 km/sec for circular motion around the earth. To get this with the smallest loss of rocket mass, we make the exhaust velocity v_e just as high as possible. How large can we possibly make v_e? This depends on the fuel, and for chemical fuels we can at best get something like 3 km/sec. To be conservative, we shall use v_e = 2 km/sec. This corresponds to gas temperatures of more than 3000° C, and only skillful nozzle design prevents much of the gas from bombarding the rocket engine casing and vaporizing it.

When $v_e = 2$ km/sec, v/v_e for a rocket that would circle the earth is $\frac{8}{2}$, or 4. Referring again to Fig. 16 – 20, we see that at this velocity ratio the mass ratio, m_r/m_o, is 0.02, or $\frac{1}{50}$. Ninety-eight percent of the original mass has to be ejected in order that the remaining 2 percent may attain a velocity of 8 km/sec. In addition, we have to get the rocket up to its orbit, forcing it up against air resistance and the gravitational pull of the earth. This means that we must eject still more mass to get the rocket in orbit. You can see why it takes tons of rocket to get kilograms of satellite into orbit.

---

*The graph and equation represent $\frac{m_r}{m_o}$ as a function of v, the relation between the fraction mass remaining and the change in velocity. They do not show anything about the time taken. The final velocity depends only on v_o, v_e, and the fraction of the original mass remaining. You can see that the time taken to burn the fuel does not affect the final velocity if you imagine that the engine is turned off for a while. During that time, the rocket coasts without changing velocity.

QUESTIONS

1.* You push a body with a force of 3 newtons for $\frac{1}{2}$ sec. What impulse do you give the body?

2.* How great is the impulse that gives an 8.00-kg mass a change in velocity of 4.00 m/sec?

3.* Suppose you throw a ball against a wall and catch it on the rebound.
(a) How many impulses were applied to the ball?
(b) Which impulse was the greatest?

4. A 3-kg body has been accelerated by a constant force of 12 newtons from 10 m/sec to 18 m/sec.
(a) What impulse was given to the body?
(b) How long was the force acting?

5. A constant force applied to a 2.0-kg object at rest moves it 4.0 m in 2.0 sec. What impulse was applied to the object?

6. A railroad freight car with a mass of 5.0×10^4 kg is rolling along a level track at 0.30 m/sec (0.7 mi/hr). A rope trails behind it.
(a) A reasonable estimate of the largest force you could apply to stop the car by pulling on the rope is 250 newtons. How long would it take you to bring the car to rest?
(b) Ten meters from the point where you start pulling, another car is standing. Will there be a crash?

7. A 10-gm rifle bullet traveling at 850 m/sec is stopped in a sandbag and is found to have penetrated 20 cm of sand.
(a) How great a force acted on the bullet? (Assume the force was constant.)
(b) How long did it take for the bullet to come to rest?

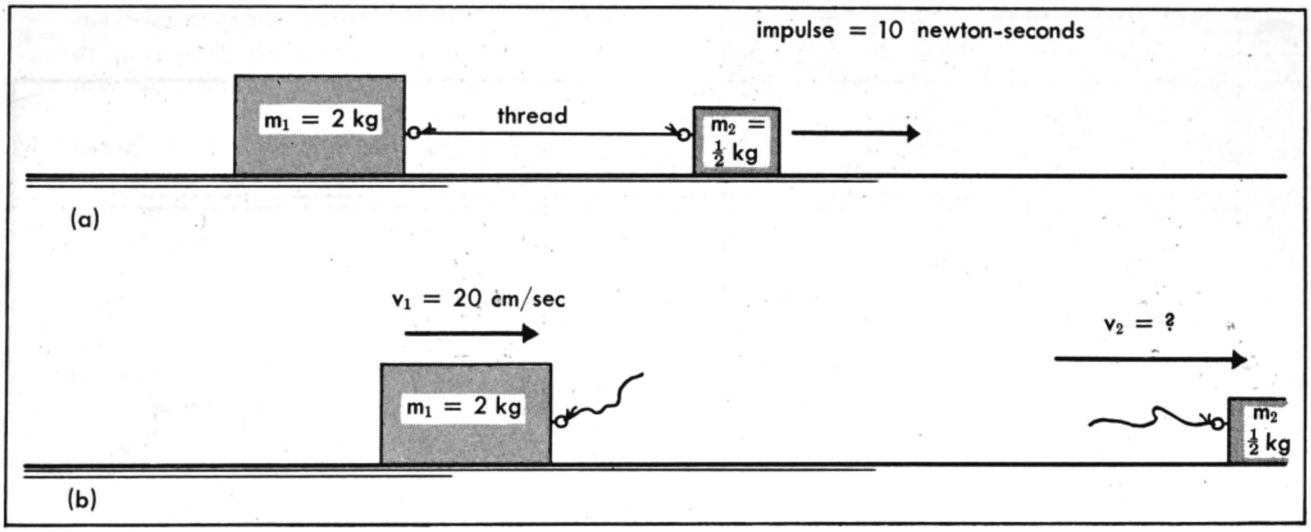
8.* What average force is necessary to stop a hammer with 25 newton-sec momentum in 0.05 sec?

9. A man is at rest in the middle of a pond on perfectly frictionless ice. How can he get himself to shore?

10. An object with a mass of 10 kg moves along a straight line at a constant velocity of 10 m/sec. A constant force then acts on the object for 4.0 sec, giving it a velocity of −2 m/sec.
(a) Calculate the impulse acting on the object.
(b) What is the magnitude and direction of the force?
(c) What is the momentum of the object before and after the force acts?

11. A body is made of two masses m_1 and m_2, 2.00 kg and 0.500 kg respectively, tied together by a light thread [Fig. 16–21 (a)]. It moves on a frictionless table starting from rest. We set it in motion by applying an impulse of 10 newton-sec to the body, but unfortunately the thread connecting m_1 and m_2 breaks during the course of the impulse [Fig. 16–21 (b)]. As a result, m_2 goes off with great speed while m_1 at the end is only moving at a speed of 0.20 meters per second.
(a) How big an impulse did m_1 receive?
(b) How big an impulse did m_2 receive?
(c) How fast is m_2 moving at the end of the impulse?

12.* An impulse is applied to a moving object at an angle of 120° with respect to its velocity vector. What is the angle between the impulse vector $\vec{F}\Delta t$ and the change in momentum vector $\Delta\vec{p}$?



16–21 For Question 11.

13.* In Fig. 16 - 12, how is diagram (b) constructed from diagram (a)?

14.* Two ice pucks initially at rest were driven apart by an explosion with velocities 5 cm/sec and −2 cm/sec. What is the ratio of their masses?

15. A 20-kg cart is moving with a velocity of 2.0 m/sec. A boy whose mass is 60 kg jumps off the cart. When he hits the ground, he is

(a) moving at the same velocity as the cart

(b) not moving relative to the ground

(c) moving with twice the initial velocity of the cart.

In each case, what is the change in velocity of the cart?

16. Two heavy frictionless carts are at rest. They are held together by a loop of string. A light spring is compressed between them (Fig. 16 - 22). When the string is burned, the spring expands from 2.0 cm to 3.0 cm, and the carts move apart. Both hit the bumpers fixed to the table at the same instant, but cart *A* moved 0.45 meter while cart *B* moved 0.87 meter. What is the ratio of:

(a) the speed of *A* to that of *B* after the interaction?

(b) their masses?

(c) the impulses applied to the carts?

(d) the accelerations of the carts while the spring pushes them apart?

17. A proton (mass 1.67×10^{-27} kg) with a speed of 1×10^7 m/sec collides with a motionless helium nucleus, and the proton bounces back with a speed of 6×10^6 m/sec. The helium nucleus moves forward with a speed of 4×10^6 m/sec after the bombardment.

(a) Can you compute the mass of the helium nucleus? If so, what is it?

(b) Can you compute the force that acted during the collision? If so, what is it?

(c) If you answered "no" to either (a) or (b), be prepared to discuss in class why you gave this answer.

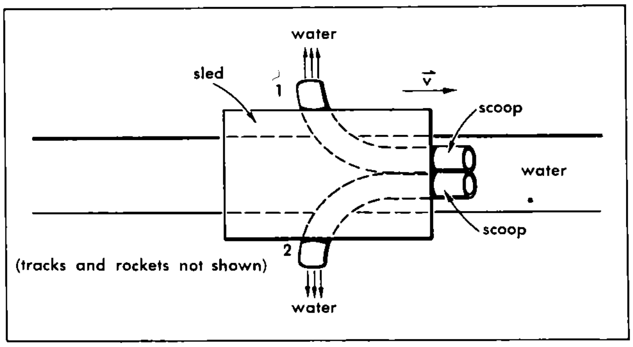
18. A stationary refrigerator car with mass 2.0×10^4 kg is rammed by a loaded gondola car with mass 3.0×10^4 kg. Before impact, the gondola car was going 1.0 m/sec. If they lock together, what is the new velocity?

19. An experimental rocket sled is slowed down by a scoop which dips into a trough of water. As shown in Fig. 16 - 23, the scoop is designed so that it ejects the water at right angles to the moving sled, and so that the water leaves the tubes with a speed equal to the speed of the sled at any instant.

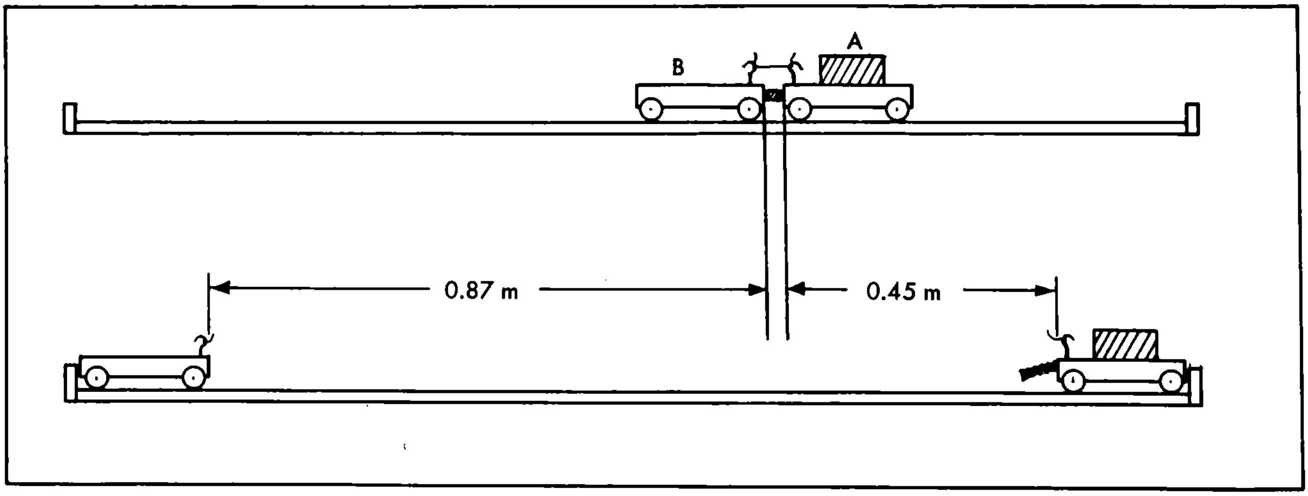
(a) What will be the change in momentum of a small mass of water *m* ejected toward either side at the instant the sled is traveling with a speed *v*?

(b) What is the change in momentum of the sled for each mass of water *m* scooped up?

(c) Is any advantage gained by ejecting half of the water from each side? Why not eject all of it from one side?



16–23 For Question 19.



16–22 For Question 16.

20. An astronaut with a total mass M (including his equipment) has been accidentally separated from his spacecraft and is at rest with respect to the ship at a distance d. His tank contains a mass m_0 of oxygen ($m_0 \ll M$) and is equipped with a nozzle through which the gas can be very rapidly ejected at an average speed of v. He must release oxygen for propulsion to get back to the spacecraft but also needs oxygen for breathing. He breathes oxygen at a rate R.

(a) If he releases an amount m of oxygen for propulsion, what speed V will he acquire? How long will it take him to reach the spacecraft?

(b) How long can he breathe on the remaining oxygen?

(c) If he is to return successfully to his craft, his breathing time, t_B, must be equal to (or greater than) his travel time t_T. What condition does this place on m?

(d) Calculate how much oxygen the astronaut may safely release for propulsion if $M = 100$ kg, $m_0 = 0.5$ kg, $d = 45$ m, $v = 50$ m/sec, and $R = 2.5 \times 10^{-4}$ kg/sec.

21.* What is the direction of motion of the center of mass of the balls in Fig. 16 – 9 after the collision?

22. While two 1-kg carts are moving along together with a velocity of 0.5 m/sec, a spring pushes them apart. One of them continues with a velocity of 0.7 m/sec, with its direction of motion unchanged.

(a) What is the velocity of the center of mass after the spring acts?

(b) What is the velocity of the other cart after it is pushed by the spring?

23. A rocket is out in free space shooting out a stream of exhaust gases and picking up speed in the opposite direction. What happens to the center of mass of all the matter—that which is ejected and that which is left in the rocket?

24. In Fig. 16 – 24, the large ball came in at the top of the picture and the little ball at the bottom. As you see, a collision took place in the middle.

(a) Draw the vectors which represent the change in the velocity of the large ball and the change in velocity of the small ball. Plot these vectors to the same scale and make sure that each one is in the right direction.

(b) Are these changes of velocity opposite in direction?

(c) Are they equal in magnitude?

(d) If their magnitudes differ, what should be their ratio?

(e) The mass of the large ball is 201 gm. What is the mass of the small ball?

25. A 2.0-kg brick with no horizontal motion is dropped on a 2.0-kg cart moving across a frictionless table at 0.40 meter/sec.

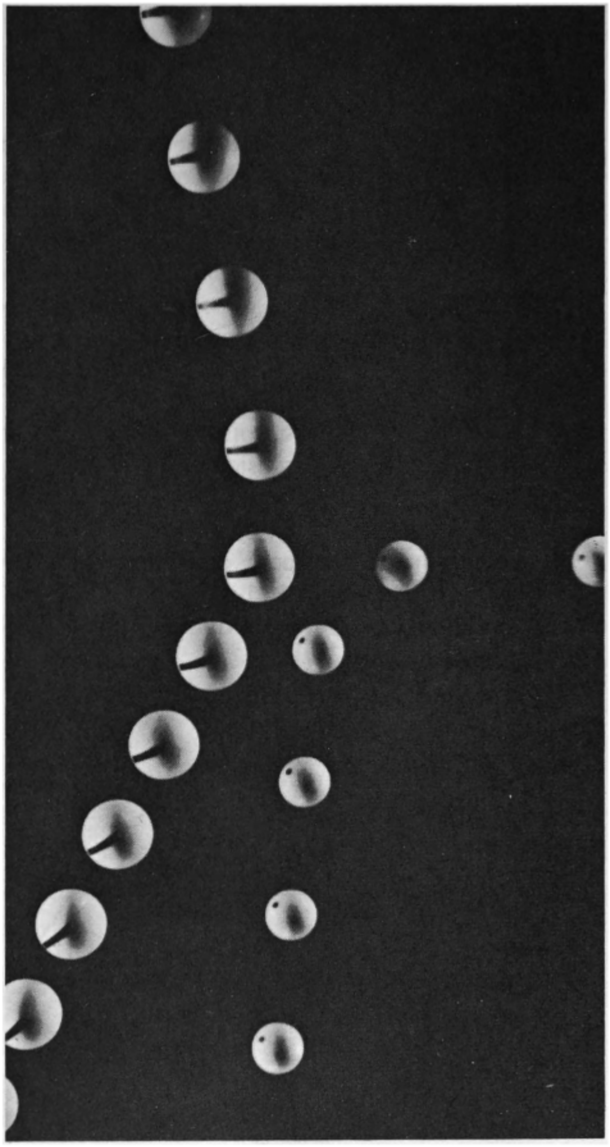
(a) What is the change in velocity of the cart?

(b) What is the velocity of the center of mass of the system composed of cart and brick before the brick is dropped?

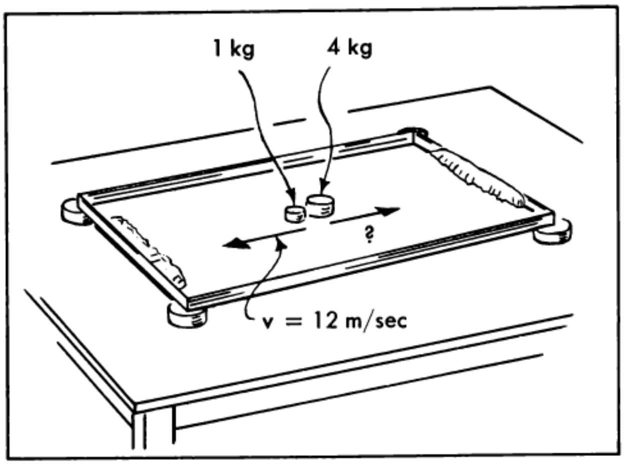
26. An explosion blows a rock into three parts. Two pieces go off at right angles to each other, a 1.0-kg piece at 12 m/sec and a 2.0-kg piece at 8.0 m/sec. The third piece flies off at 40 m/sec.

(a) Draw a diagram to show the direction in which the third piece goes.

(b) What is its mass?



16–24 For Question 24.



16–25 For Question 27.

27. The system shown in Fig. 16 – 25 consists of a 5.0-kg frame with a 1.0-kg and a 4.0-kg mass in the middle. At either end of the frame are stops made of putty. The two masses and the frame are mounted on Dry Ice pucks. (The mass of the Dry Ice is included in the values given above.) The center of mass of the whole system is at the center of the frame. An explosion pushes the two masses apart. The 1.0-kg mass goes off with a speed of 12 m/sec. Eventually each mass is trapped by the putty.

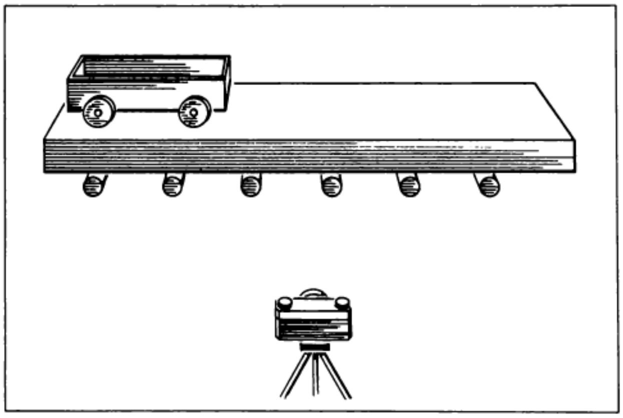
   (a) What is the speed of the 4.0-kg mass immediately after the explosion?

   (b) Where is the center of mass of the whole system after 1 sec? After 2 sec?

   (c) What is the velocity of the frame and the attached masses after 100 sec?

   (d) Describe the motion qualitatively from the time of explosion to 100 sec.

   (e) How far does the frame move?



16–26 For Question 28.

28. A 20-kg car stands at rest on an 80-kg platform as shown in Fig. 16 – 26. The car can be driven along the platform by a motor and the platform, mounted on rollers, is free to move along the laboratory floor substantially without friction. A stroboscopic camera records any motions that occur.

   When the motor is turned on, a study of the photograph shows that over a period of 3.0 sec the platform acquired a velocity of 0.30 m/sec.

   (a) What does the same photograph show for the velocity of the car at the end of the 3.0-sec period?

   (b) With what force did the wheels of the car push against the platform?

   (c) What was the relative velocity of the car along the platform at the end of the 3.0 sec?

   (d) If you know only the relative motion of the car along the platform, what force would you calculate was needed to give the car this velocity in 3.0 sec? Can you explain why the force you calculate here is different from the one you found in part (b)?

29. A body of mass m_1 moves with velocity $\vec{\mathbf{v}}_1$ toward a body of mass m_2 which is at rest. The total momentum of the two bodies is $m_1\vec{\mathbf{v}}_1$. This is the situation as described in the frame of reference in which the second body is at rest, often called the *laboratory frame*.

   (a) Consider another frame which moves with a velocity $\vec{\mathbf{u}}$ with respect to the laboratory frame. What must $\vec{\mathbf{u}}$ be so that the total momentum of the two bodies in this frame (which is called the *center-of-mass frame*) will be zero?

   (b) The two bodies collide, not necessarily head-on. Can you determine the angle between their momentum vectors in the laboratory frame? Can you determine the angle between their momentum vectors in the center-of-mass frame?

30.* What happens to the momentum of a car when it comes to a stop?

31. A double star consists of two large masses that attract each other gravitationally. By observing the motion of both masses, we can see that they rotate around each other.

   (a) What do you think happens to the momentum of each of the masses in a double star as time goes on? Explain your answer.

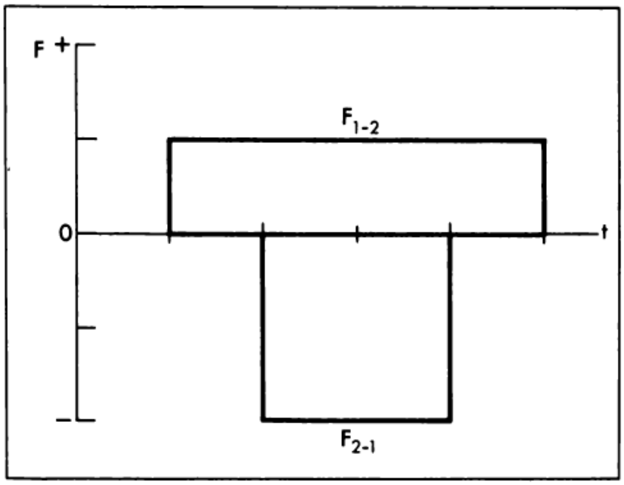
   (b) When observed carefully, the bright star Sirius seems to wobble about, instead of having a uniform motion of the center of mass. From this and other evidence astronomers believe that Sirius has a dark companion. It is really a double star. How does this explain the peculiar observed motion?

32.* Consider a collision between two particles in which light is emitted in a given direction. Since light exerts a pressure, it must have momentum. If the overall momentum (including the momentum of the light) is conserved, are the forces of interaction between the two particles equal and opposite?

33. Figure 16 – 27 shows the hypothetical force-time graphs between two bodies; F_{1-2} and F_{2-1} are the forces with which the two bodies act on each other.

(a) Is the total final momentum different from the total initial momentum?

(b) Is momentum conserved throughout the collision?



16–27 For Question 33.

34. (a) Use the graph in Fig. 16 – 20 to plot on 3 cycle (3 orders of magnitude) graph paper the ratio $\frac{m}{m_0}$ as a function of v for a rocket with a constant continuous exhaust velocity of 2 km/sec.

(b) From the graph in part (a) determine the fraction of the original mass of the rocket that remains when the speed of the rocket reaches 12 km/sec.

17

Work and Kinetic Energy

ALTHOUGH energy may be hard to define precisely, it is familiar to all of us. The energy we need to live and work is supplied to us through the food we eat. We need other fuels too: wood, natural gas, oil, gasoline, etc., for cooking, warming our homes, or moving our cars. For thousands of years, man used the energy released by burning fuels for such purposes as keeping warm and, on a small scale, as an aid to his arts and crafts. Until only recently all heavy work was done by muscles, either human or animal. Since the Industrial Revolution, however, man has been using more and more energy from other sources to replace muscular effort. Burning fuels now supply most of the energy needed to operate our farm machinery, run our factories, and drive our trains, autos, ships, and airplanes. The energy taken from fuels is an absolute necessity to our present technical society. One very good measure of the technological progress and material well-being of a country is the average amount of energy taken from fuel for each citizen during the year.

In everyday language, we often speak of energy in the following way: we say we are "full of energy" if we get up in the morning eager and ready to tackle a job or take on a challenge. But when we have been active for some time, we get tired and say we have "lost our energy." A good night's rest and some nourishing food will reestablish our capacity to do a job. This common use of the word "energy" is closely related to the technical idea of energy we wish to develop. It points up the close connection between energy and the fuels which enable both muscles and machines to do useful jobs. For the moment we shall say that energy is the essential thing involved in jobs—not the creation of energy but its transfer from one form to another.

17–1 ENERGY TRANSFER

Jobs are done by the transfer of energy from one form to another or from one place to another—for example, from energy in fuel to energy in a speeding train; from heat in a boiler to heat in a room. Fuels, then, hold energy in stored form, ready for use. A given amount of fuel contains a given amount of energy. You will see later that energy is never created or destroyed, but only transferred (from one form to another or from one place to another). It is somewhat like cash, which changes hands only when someone buys something.

Consider what we could do with the energy stored in some gasoline. We might burn the gasoline in an automobile engine and use the energy that can be transferred to the wheels to run the car up an incline. Or we might burn the gasoline under a steam boiler. There energy is transferred into the steam. The steam can expand and move a piston which can raise a load against gravity. The load may then be hitched to the axle of a car by a rope passing over a pulley; if the load is then allowed to fall, it can pull the car up the incline. Energy has again been transferred; at least some of the energy in the fuel has gone into raising the car, but by a more roundabout method than before. We must now seek some detailed accounting procedures to measure these transfers of energy through various jobs; when we learn to account for all such transfers of energy, we shall find that energy is not created or lost; energy is conserved.

17–2 WORK: A MEASURE OF ENERGY TRANSFER

When energy is transferred in some fuel-using job, we can measure the transfer, the cost of the job, either by (a) the number of unit jobs (e.g., number of lawns mowed, or the number of houses heated), or by (b) the amount of fuel used (e.g., the number of gallons of gasoline).

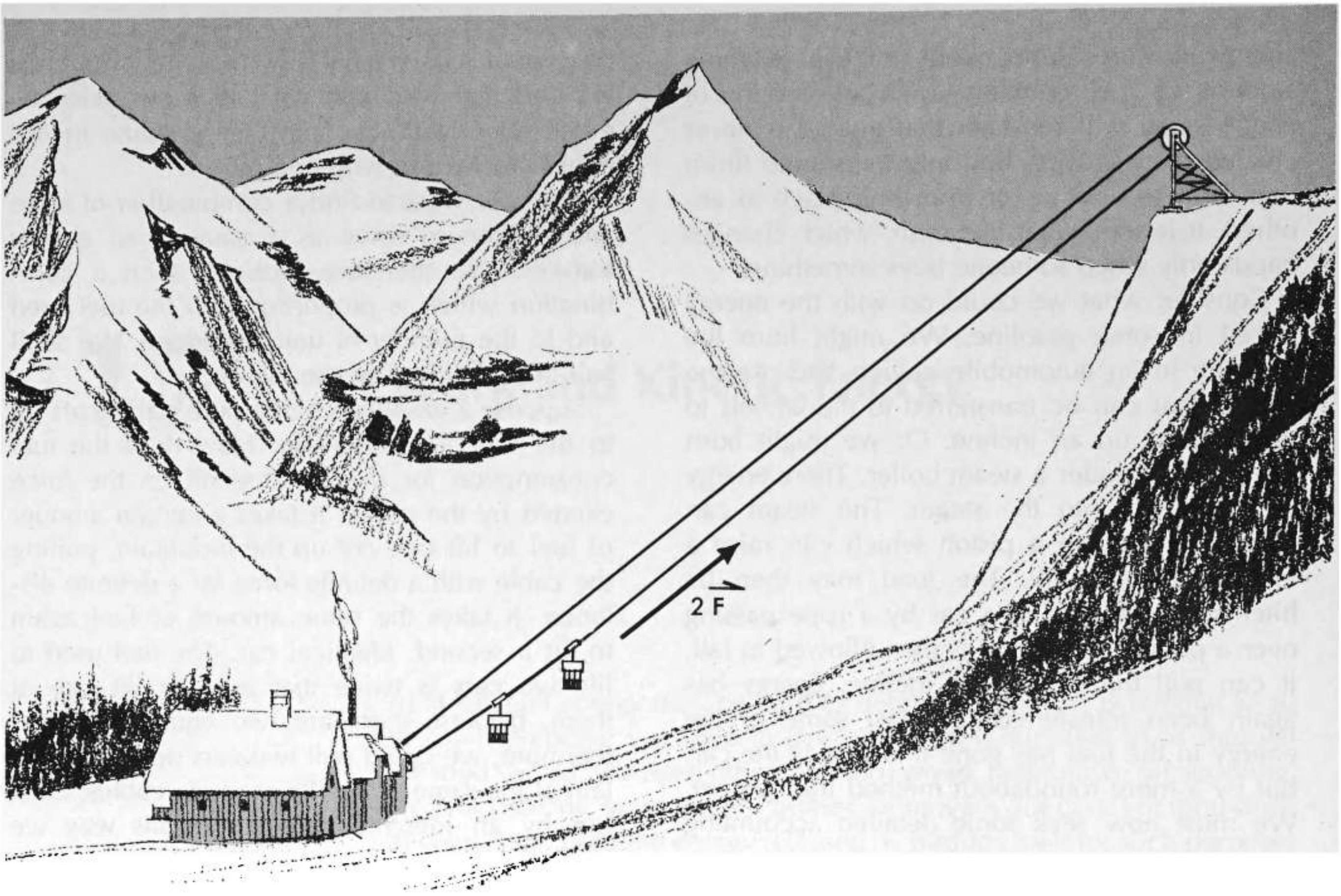
In the jobs just described, forces are exerted and things move. We say that work has been done in those transfers of energy, and we shall now develop the idea of work as a precise and reliable measure of energy transfer. For the moment, we shall use *work* to mean the measure of the job done or of the fuel used. Notice that these two measures of energy transfer agree with each other. Think of any job: you will see that the number of unit jobs in it and the amount of fuel used are proportional. To do two equal jobs we need twice as much fuel as for one. For instance, we can mow twice as much lawn with two pints of gasoline in the mower as we can with one pint.

We now want to find a combination of force and motion to serve as a measure of energy transfer. We therefore look for such a combination which is proportional to the fuel used and to the number of unit jobs done. We shall call this combination the "work."

Suppose a gasoline engine pulls cable cars up to the top of a mountain. How does the fuel consumption for the job depend on the *force* exerted by the cable? It takes a certain amount of fuel to lift one car up the mountain, pulling the cable with a definite force for a definite distance. It takes the same amount of fuel again to lift a second, identical car. The fuel used to lift two cars is twice that used to lift one of them, because there are two equal jobs. Furthermore, we could pull two cars up the mountain at the same time with separate cables, each run by an identical engine. In this way we could lift them simultaneously, exerting twice the force and burning twice the fuel needed to lift one. We could tie the two cables together, side by side, and that would make no difference to the fuel burned or to the job done. We could even replace this double cable by a single cable supporting both cars, and use one engine to do the job—but that engine would have to burn fuel faster. The fuel used to lift the two cars would still be twice that needed to lift one. The fuel needed is doubled when the force exerted is doubled. So we say work, which we are going to use as the measure of energy transfer, is proportional to the force exerted (Fig. 17 – 1).

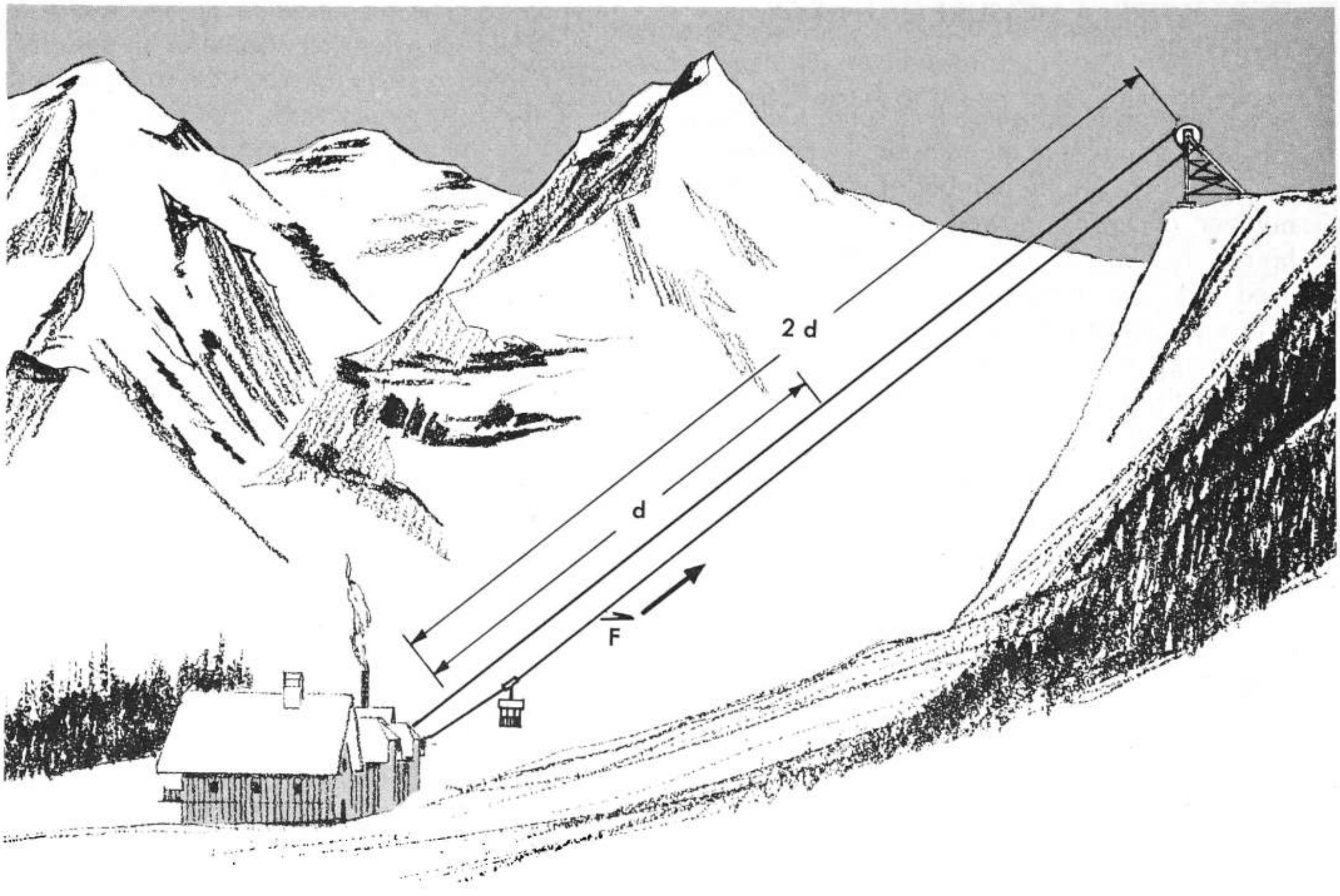
Now consider how the fuel used depends on the distance a car moves up the mountain (Fig. 17 – 2). Lift the car up halfway, then up to the top. To lift the car up the second half of the mountain, the engine exerts the same force and moves the same length of cable as it does to lift the car up the first half. The same job is done twice; so it takes twice as much fuel to move a car all the way as it does to move it halfway. The fuel used is doubled when the distance is doubled; so we say work is also proportional to distance moved. Combining these results we define work as

force × distance moved



17–1 **Pulling two identical cable cars up the mountain requires twice the fuel needed to haul up one; the force is doubled also. Thus we see that the work must be proportional to the force.**

17–2 **It takes twice as much fuel to pull the cable car to the top of the mountain as it does to pull it up halfway. Work is therefore proportional to the distance moved.**



From a study of many other examples we have concluded that "work" thus defined is a good way of measuring energy transfer.

The unit of work must be *unit force* times *unit distance: one newton* times *one meter,* or *one newton-meter.* We call this unit one joule after the English physicist James P. Joule (1818–1889), who did some of the first careful experiments in measuring energy transfers to establish the conservation of energy. If we move a body 10 meters while pushing with a force of 2 newtons in the direction of motion we do 20 joules of work.

When we lift a car up the mountain rapidly, we use practically the same amount of fuel as when we lift it slowly. Time has nothing to do with the amount of work in the job. As an extreme case, we can do half the job today and half tomorrow, and still use the same amount of fuel.

If we put the brakes on and turn off the engine, the car stops and thereafter no energy is transferred to it from the fuel. Although a force is still exerted to hold the car there, no work is done because there is no distance moved.

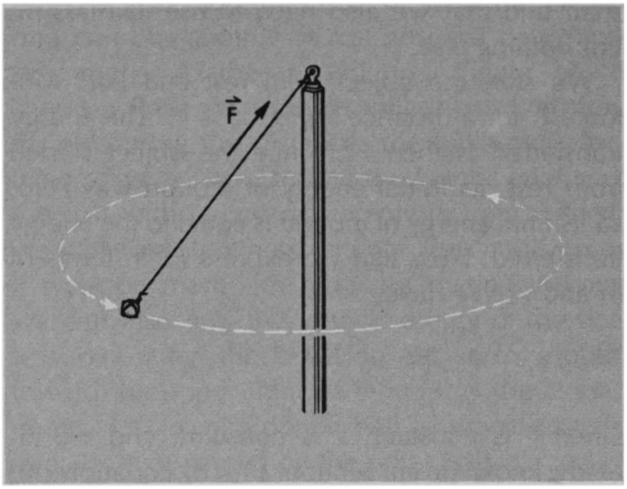
But energy can be taken from fuel without doing a useful job. We can disengage the cable from the engine and let the engine run, burning fuel even though it is not pulling the car up the mountain. The fuel gives energy to the motion of the pistons, moving them through distances with forces opposite to those of friction. In this way, energy goes into heating the engine. Energy is still being transferred from fuel to other forms and places though none is used to do the job of lifting the car. Even when the engine is pulling the cable car up the mountain, some of the fuel's energy does these other things. The amount of fuel actually burned may differ from trip to trip—for example, if the pistons are poorly lubricated on one trip, more fuel will be used. Correcting for these effects, however, we find that the fuel used for the job of lifting the cable car is always the same.

17–3 MORE ABOUT THE DEFINITION OF WORK

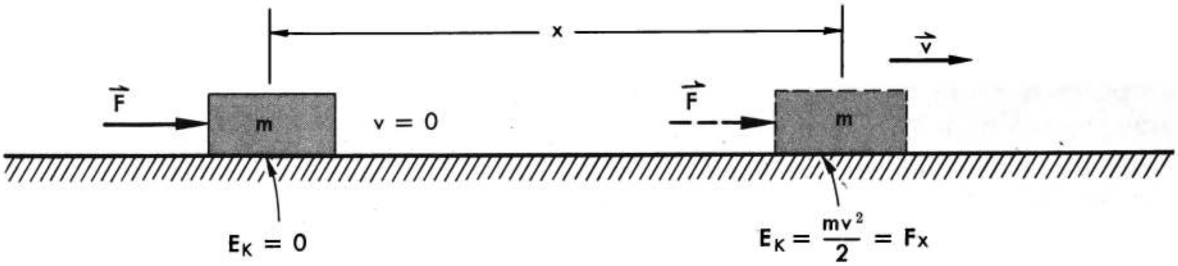
We developed the definition of work as force times distance from the example of a cable pulling a car which moved along the direction of the pull. When the force is in the direction of motion, there is a transfer of energy; work is done. On the other hand, there is no energy transfer when we exert a force on a body perpendicular to the direction of its motion. For example, consider a stone whirling around on the end of a string which is tied to a rigid post (Fig. 17–3). The stone moves around and around without changing speed. Time and time again it comes to the same place moving at the same speed; and its energy, also, is unchanged. At the other end of the string, the post remains motionless. It burns no fuel, and transfers no energy to the string. Therefore, the string neither gives nor receives energy; so its pull does no work.

What happens if we pull with a force at some other angle to the motion? Then the force has a component in the direction of motion, and this component speeds up the stone. After that, because it is moving faster, the stone can do more work pulling something as it comes to rest. It therefore has more energy. The component of force in the direction of motion has done work on the stone, transferring energy of motion to it.

Briefly, the component of any force perpendicular to the motion does no work. *The work done on an object is equal to the magnitude of the component of the force along the direction of motion times the distance moved.*



17–3 A stone at the end of a string is whirling around a post. The string is attached to a frictionless bearing on the post. No work is done transferring energy to the stone, because there is no component of force in the direction of the stone's motion.



17–4 A force $\vec{F}$ is exerted on a mass m, which moves a distance x starting from rest. The energy transferred to m is Fx, equal to $mv^2/2$, which we call the kinetic energy of the body.

17–4 KINETIC ENERGY

Our definition of work in terms of the force and the distance moved has been designed to agree with the idea that equal amounts of fuel will supply equal amounts of energy. Does this definition of work also enable us to say how much energy a moving body has? The value of our definition of work depends on the answers to questions like this.

Suppose we apply a force to a body on a frictionless table. The body accelerates, gaining speed. While the force acts, we do work transferring energy to the body. We calculate the work done on the body as it moves a measured distance under the action of the force. This work is the energy transferred, and it gives us an expression for the energy of motion of the body. This energy is called kinetic energy. We want to find an expression for it *in terms of the motion,* using the object's speed v. We shall find that we also need to use its mass m, but nothing else.

We start the object from rest and pull with force $\vec{F}$ for a distance x (Fig. 17 – 4). The energy transferred is then Fx. Since the object started from rest, its initial energy of motion was zero, so its final energy of motion is equal to the energy transferred. How can we express Fx in terms of m and v? We have

$$F = ma$$

so

$$Fx = max$$

Since F is constant, a is constant; and we already know [from Section 11 – 8, Equation (3)] that

$$v^2 = 2ax$$

or

$$ax = v^2/2$$

Therefore,

$$Fx = max = mv^2/2$$

The work $W = Fx$ done accelerating the mass m from rest is equal to the quantity $mv^2/2$, or in other words, $mv^2/2$ is equal to the energy transferred to the body in setting it in motion. We call it E_K, the *kinetic energy* of the body: that is,

$$E_K = \frac{mv^2}{2}$$

Notice that the kinetic energy $mv^2/2$ is defined purely in terms of the mass and its motion. It shows no trace of the force used to give the body this energy, and no trace of the distance moved. It does not depend on the way in which the energy was transferred to the mass. If we know the mass, we can evaluate the kinetic energy by observing how fast the body moves. In other words, our definition of E_K is in terms of the state of motion of the mass without reference to its history. Whenever the mass is in the same state of motion, $mv^2/2$ is the same. The body has the same kinetic energy.

We can easily see that the work required to give a body that starts from rest a certain $mv^2/2$, a certain kinetic energy, is always the same. Try using double the force for half the distance; you will get the same $mv^2/2$ for the same work. Suppose that the force changes to a new value when the body is already moving with speed v_1, and the work $W_1 = mv_1^2/2$ has been done. The body now is accelerated with a different constant acceleration a', and from Section 11 – 8 we know that after an additional distance Δx its speed v is given by

$$v^2 = v_1^2 + 2a'\Delta x$$

On multiplying by $m/2$ we see that $mv^2/2$ at the end is

$$\frac{mv^2}{2} = \frac{mv_1^2}{2} + ma'\Delta x$$

Since ma' is the new force F',

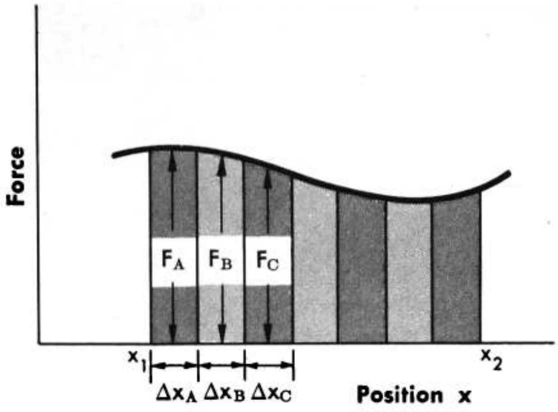
$$ma'\Delta x = F'\Delta x$$

This is the additional work ΔW; and the equation above shows that it increases the kinetic energy from $mv_1^2/2$ to $mv^2/2$. The total work is therefore the sum of the work W_1 (which gives transfer of energy from zero to $mv_1^2/2$) and the additional work ΔW (which gives the change from $mv_1^2/2$ to $mv^2/2$). The total work from rest to speed v, this time given by $W_1 + \Delta W$, again gives $mv^2/2$.

If the force continually changes during the motion (as in Fig. 17 – 5, for instance), we can always take small intervals during which its value is nearly constant. In each interval, as we have seen, the work will give the *change* in $mv^2/2$. So the total work will always give the total change in $mv^2/2$. Indeed, even if a force points in some direction other than that of the motion, we still get the same result. Only the component of force along the motion enters the work, and only the component along the motion changes the speed. Consequently, our definition of kinetic energy as $mv^2/2$ is such that the same amount of work always gives the same change of kinetic energy.

If a mass m is moving with speed v, its kinetic energy is $mv^2/2$. This amount of energy had to be transferred to it to accelerate it from rest to the speed v. When a force acts on the mass to slow it down, the kinetic energy decreases. Energy is transferred from kinetic energy of the mass to energy in the system that exerts the retarding force. The work $F_x\Delta x$, when the force F_x retards the motion, measures the energy coming out of the moving mass as it slows down. By using the same reasoning as above, we see that if the body has $mv^2/2$ kinetic energy, it can do just that amount of work before coming to rest, no matter how it is slowed down.

In summary, when work is done by a force acting on a mass in the direction of its motion, the kinetic energy of the mass is increased. The work measures the transfer of energy from the outside into kinetic energy of the mass. When a force opposes the motion, the transfer of energy is in the opposite direction and the work measures the transfer of kinetic energy of the mass into energy of something else. We can express the kinetic energy of a mass m moving with a speed v directly in terms of its motion as $mv^2/2$.

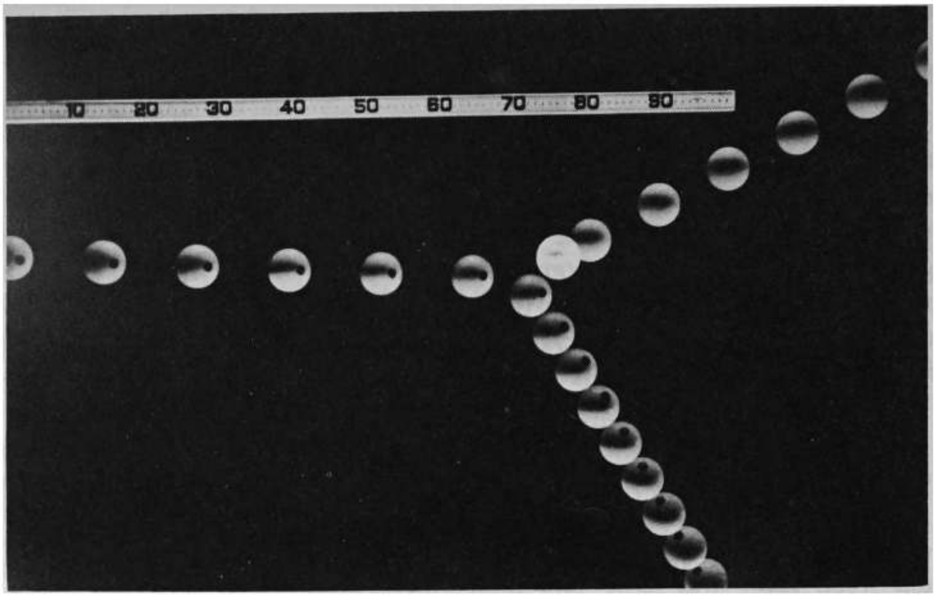


17–5 If a force changes in magnitude but is always in the direction of a body's motion, the total work it does can be calculated by taking the sum of the areas of the many small rectangles with base Δx and height F. When the body moves from x_1 to x_2, the total work is represented by the shaded area. It gives the total change in $mv^2/2$.

17–5 THE TRANSFER OF KINETIC ENERGY FROM ONE MASS TO ANOTHER

So far we have examined what happens when one object gains or loses kinetic energy. This gain or loss is measured by the work done in transferring the energy to or from the object. We believe that energy transferred to the object comes from someplace, and energy transferred from it goes somewhere. For example, when one ball hits another at rest, the first ball transfers energy to the second one. Figure 17 – 6 shows a flash photo of a billiard-ball collision. By measuring the distances the balls move between flashes, we can find the kinetic energy of the ball with the dot as it came in from the left; we can find its kinetic energy after collision as it went off more slowly to the lower right, and we can also find the kinetic energy of the dotless ball after the collision set it in motion toward the upper right. We find that the kinetic energy lost by the dotted ball is almost exactly equal to that gained by the other ball.

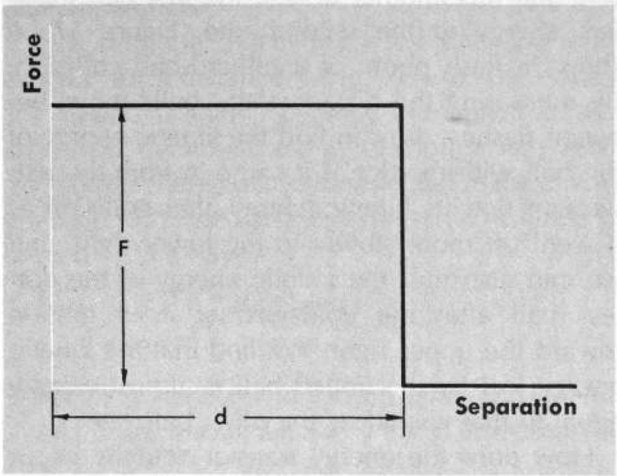
How does the energy transfer actually occur in a billiard-ball collision? The balls approach, exerting no force on each other until they are very close, and we say they are "in contact." Then repulsion sets in: each ball pushes the



17–6 A multiple-flash photograph of a collision between two billiard balls, each of mass 173 gm. The interval between flashes was 1/30 sec. The dotted ball came in from the left and struck the unmarked ball at rest. The unmarked ball appears whiter where it was at rest and was photographed during several flashes.

other. These forces change in magnitude, growing and decreasing again as the balls distort each other and finally push each other away. Because the forces are changing, a billiard-ball collision is difficult to analyze in detail. We shall be able to understand the transfer better if we start with an easier example in which the force has a simpler behavior. After we see the details in this simple case we can return to billiard balls.

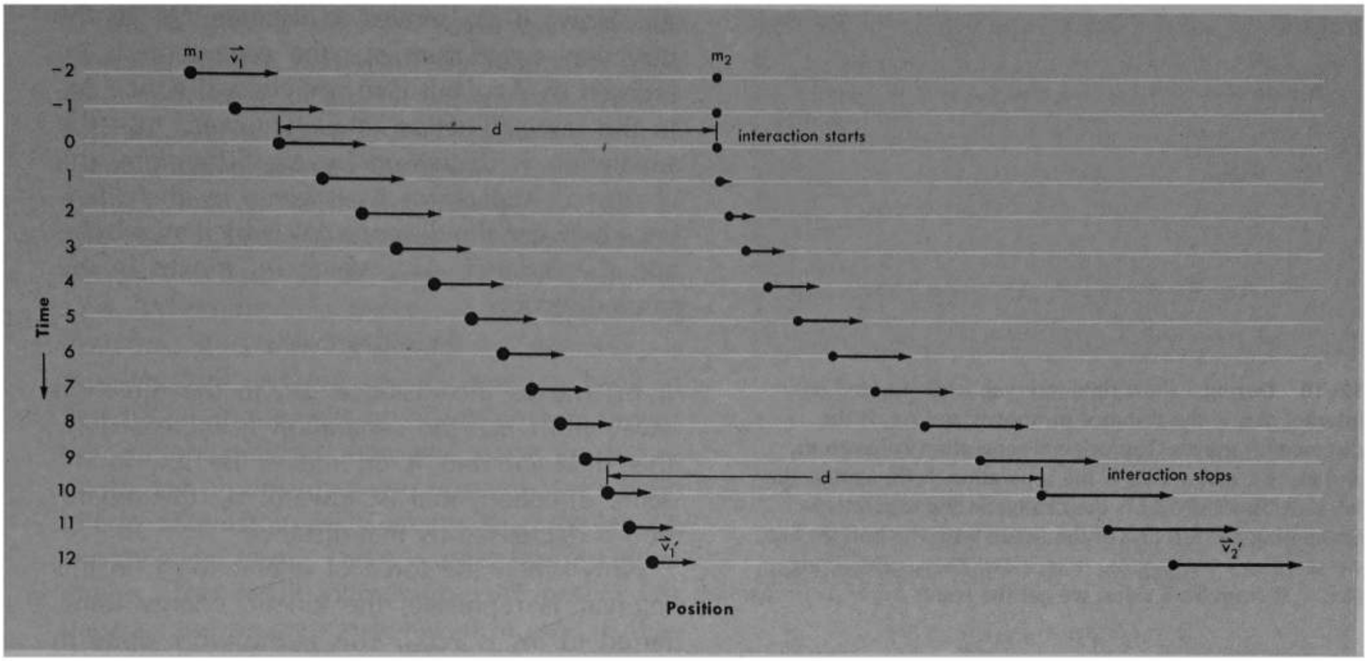
To make a simple preliminary study of collisions and energy we shall study an artificial case in which two colliding bodies exert no force on each other until they get within a distance d and then exert a constant repulsion of magnitude F on each other as long as the separation is less than d (Fig. 17 – 7).



17–7 The graph of the magnitude of the interaction forces between two bodies. The force is zero when the separation s is more than d, and a constant value F when the masses are closer together than d.

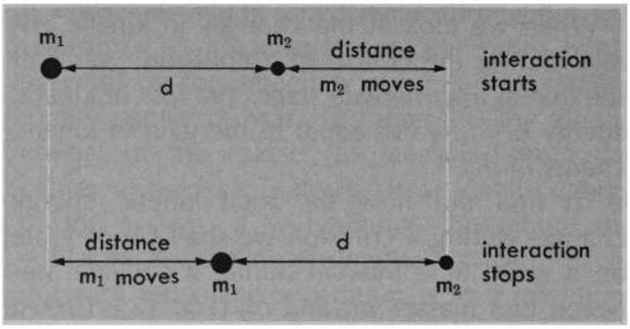
Suppose the collision takes place along a line (Fig. 17 – 8). Moving along this line with velocity v_1, the mass m_1 approaches the mass m_2 which is at rest. Nothing happens until m_1 gets to the distance d away from m_2. From then on, however, m_2 is pushed forward with force F, and m_1 is pushed backward just as hard. Consequently m_2 speeds up while m_1 slows down. Because m_1 is moving and m_2 standing still when the repulsion sets in, the masses continue to get closer together for a while. But soon they reach a stage when they are as close as they will get and are moving along with the same velocity. (If m_1 were still going faster it would still be gaining on m_2; and if m_2 were going faster, the masses would be separating.) Still pushed apart by F, m_2 continues to speed up and m_1 to slow down. The masses therefore separate until eventually they are d apart again. Then, because they are getting farther apart, the force falls to zero, and they continue their further motion without changes in velocity. The collision is over, with m_1 and m_2 now moving with definite energies that no longer change.

For this collision with a steplike force, we can calculate the details of the motion of each mass. This we have done, using Newton's law, for masses such that $m_1 = 3m_2$; and we constructed Fig. 17 – 8 in this way. But we do not need to go through the details of calculating the motion to find out whether the transfer of kinetic energy



17–8 The positions and velocities of two masses m_1 and m_2 as they interact along a line, with the force shown in Fig. 17–7. The masses are shown at equal time intervals. Mass m_1 comes in from the left. When the distance between masses is less than d, m_1 slows down and m_2 speeds up from rest. After 10 time intervals, in this particular case, the separation is again greater than d; the force drops to zero; and the two masses move apart with constant velocities. The gain in kinetic energy of m_2 over the whole interaction is equal to the loss of kinetic energy of m_1.

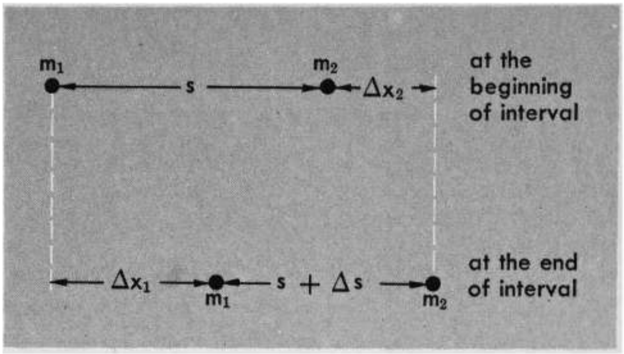
from m_1 is the same as the transfer of kinetic energy to m_2. Because the force has the constant value F during the interaction, the energy transferred into kinetic energy of m_2 is F times the distance that m_2 moves during the collision. Also, the kinetic energy transferred from m_1 is F times the distance m_1 moves during the collision (m_1 loses kinetic energy because the force on it opposes the motion.) We can show that the kinetic energy lost by m_1 is just the amount gained by m_2 if we show that m_1 and m_2 move the same distance during the collision. To see that, look at Fig. 17 – 9. There we show the positions of the masses when the interaction starts and when it stops. Notice that d plus the distance m_2 moves during the interaction (top line) is equal to d plus the distance m_1 moves (bottom line). Consequently the distances moved by m_1 and by m_2 are the same during the complete interaction. The force times distance moved during the collision is therefore equal and opposite for the two masses; and as we anticipated, the kinetic energy from m_1 just goes into kinetic energy of m_2.



17–9 The positions of m_1 and m_2 at the beginning (top line) and the end (bottom line) of the interaction in Fig. 17–8. The scale here has been reduced. Comparing the two lines, d plus the distance m_2 moves is equal to d plus the distance m_1 moves. Thus m_1 and m_2 move the same distance during the interaction.

17–6 ANOTHER LOOK AT THE SIMPLE COLLISION

There is a slightly different way of looking at this special example that will make it easier to see what happens in a real collision when the pattern of force is more complicated. In order



17–10 **During a short time interval, while m_1 and m_2 interact, Δx_1 is the distance m_1 moves and Δx_2 is the distance m_2 moves. The original separation between m_1 and m_2 is s, and $s + \Delta s$ is the separation at the end of the interval. Therefore Δs is the change in the separation. Comparing the top line of the figure with the bottom line, we see that $s + \Delta x_2 = \Delta x_1 + (s + \Delta s)$. Then, subtracting $(\Delta x_1 + s)$ from both sides, we get the result $\Delta x_2 - \Delta x_1 = \Delta s$.**

to study real collisions we wish to find what kind of forces must be acting if all the kinetic energy lost by one mass is gained by the other. To find the nature of these forces we shall examine the sum of the two kinetic energies directly rather than each one individually. If the sum at the end is the same as at the beginning, what is lost by one mass must be gained by the other.

When we look at the changes in kinetic energy during the course of a collision, we shall see that at *intermediate stages* the loss of kinetic energy of m_1 is not equal to the gain of kinetic energy of m_2.

To find out how the total kinetic energy changes during a collision we shall concentrate on a small time interval during a collision between two masses m_1 and m_2 (Fig. 17 – 10). At the beginning of this time interval, we shall call the separation between the masses s. At the end of this time interval the separation is $s + \Delta s$. Now, if m_2 moves a distance Δx_2 in the direction away from m_1, the separation is increased by Δx_2; but if m_1 moves a distance Δx_1 in the same direction—that is, toward m_2—the separation is decreased by Δx_1. Therefore, the change in separation Δs is given by the difference between the distance Δx_2, which m_2 moves, and the distance Δx_1, which m_1 moves in the same direction:

$$\Delta s = \Delta x_2 - \Delta x_1$$

If m_2 moves the distance Δx_2 in the direction away from m_1, the separation is increased by this distance; but if m_1 moves by Δx_1 in the same direction, that is, toward m_2, the separation is decreased by that distance.

Now when the force of interaction F in this interval is repulsive, the kinetic energy transferred to m_2 is $F\Delta x_2$. This is the work done in accelerating m_2. On the other hand, the kinetic energy taken from the motion of m_1 is $F\Delta x_1$. This energy is transferred from the motion of m_1 because the force on m_1 is in the opposite direction from the motion Δx_1. Consequently, the change in the total kinetic energy in the interval when these motions take place is given by

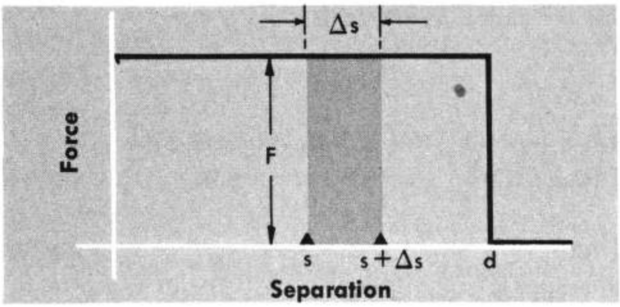
$$\Delta E_K(\text{tot}) = F\Delta x_2 - F\Delta x_1 = F(\Delta x_2 - \Delta x_1)$$

Since $\Delta x_2 - \Delta x_1 = \Delta s$, we can write the change in the total kinetic energy as

$$\Delta E_K(\text{tot}) = F\Delta s$$

For example, in Fig. 17 – 11 as the separation changes from s to $s + \Delta s$, the total kinetic energy changes by the amount given by the shaded area, F high and Δs wide.

The result $\Delta E_K(\text{tot}) = F\Delta s$ is just what we need. It shows that the change in the total kinetic energy in any interval depends only on the force of interaction and the change of separation that takes place. These were the only quantities that entered our calculations. We never have to say where the masses are with respect to any other objects. Also, it does not matter what the interaction force may be at some other separation. Consequently, the force that acts at other separations need not have the same value that it has in the interval we just considered. It is for this reason that we shall be able to apply our results to collisions in which the force of interaction depends in any way on the separation alone. We shall see how



17–11 **When the separation changes from s to s + Δs, the change in the total kinetic energy is given by the shaded area.**

to use $\Delta E_K(\text{tot}) = F\Delta s$ for such forces in the next section. We shall use this equation to show the conservation of kinetic energy in the simple collision we studied in the last section.

For a collision with the force shown in Fig. 17 – 11 (that is, F at all separations less than d, and zero at larger separations), the total kinetic energy decreases at the beginning of the collision. When m_1 first arrives at the distance d from m_2, it is moving toward m_2 while m_2 is standing still. As we see by looking back at Fig. 17 – 8, in the next interval of time m_1 moves considerably farther than m_2. Since the forces on m_1 and m_2 are equal and opposite, the kinetic energy taken from m_1 is greater than the kinetic energy given to m_2. There is a net loss in kinetic energy. This result follows from $\Delta E_K(\text{tot}) = F\Delta s$ because the separation between m_1 and m_2 has decreased, and so $F\Delta s$ is negative.

As long as m_1 is coming closer to m_2, kinetic energy continues to disappear. Eventually, however, the total kinetic energy is regenerated. While m_1 continues to slow down, m_2 picks up speed so that finally the separation increases again. When the separation becomes d once more, the value of Δs, measured from the beginning of the collision, is zero. Consequently $\Delta E_K = F\Delta s = 0$. The kinetic energy which was lost as the two bodies approached each other has been regained. After that the masses separate still farther but there is no force acting on them. Consequently E_K stays the same. The collision is over and the kinetic energy remains at its final value, the same value that it had before the collision began.

We can look at the kinetic energy transfer in this collision in yet another way. We learned in the last chapter that the center of mass moves steadily through a collision as if nothing had happened. Also, as far as momentum was concerned, the center of mass moved as if all the mass of the system were concentrated there, and the total momentum relative to the center of mass was zero. Is there a similar behavior for the center of mass where kinetic energy is concerned?

At any instant in the collision of Fig. 17 – 8, the masses m_1 and m_2 are moving along a line with velocities v_1 and v_2. The center of mass moves along the same line with velocity v_c. The two masses have velocities V_1 and V_2 relative to the center of mass, also along this line. The velocity v_1 is just the velocity of m_1 relative to the center of mass plus the velocity of the center of mass: $v_1 = V_1 + v_c$. Similarly, for v_2 we have: $v_2 = V_2 + v_c$ (remember that velocities along a line add like positive and negative numbers). The total kinetic energy of the masses is then

$$\begin{aligned} E_K &= \tfrac{1}{2}m_1v_1^2 + \tfrac{1}{2}m_2v_2^2 \\ &= \tfrac{1}{2}m_1(V_1 + v_c)^2 + \tfrac{1}{2}m_2(V_2 + v_c)^2 \\ &= \tfrac{1}{2}m_1V_1^2 + \tfrac{1}{2}m_2V_2^2 \\ &+ (m_1V_1 + m_2V_2)v_c + \tfrac{1}{2}(m_1 + m_2)v_c^2 \end{aligned}$$

Now, we found in Section 16 – 5 that $m_1V_1 = -m_2V_2$: the total momentum relative to the center of mass was zero. Therefore

$$(m_1V_1 + m_2V_2) = 0$$

and

$$E_K = (\tfrac{1}{2}m_1V_1^2 + \tfrac{1}{2}m_2V_2^2) + \tfrac{1}{2}(m_1 + m_2)v_c^2$$

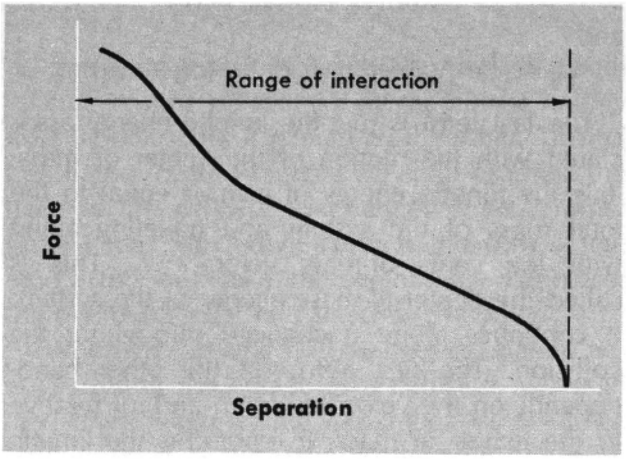
The last term is just the kinetic energy associated with the motion of the center of mass. It is the kinetic energy of a mass equal to the total mass of the system and traveling along with the center-of-mass velocity v_c. This is called the center-of-mass energy of the system. It continues along unchanged throughout the collision. The first term, on the other hand, depends on the velocities of m_1 and m_2 relative to the center of mass; it represents the kinetic energy of the masses as measured from the center of mass. It is the *internal* kinetic energy of the system. It will change throughout the collision, decreasing during the first part, reaching zero when the separation is at a minimum ($v_1 = v_2 = v_c$ at that instant so $V_1 = V_2 = 0$), and then increasing back to its original value by the end of the elastic interaction.

This separation of the kinetic energy into two parts is not limited to two bodies moving along a straight line; it can be proved for any number of bodies moving in all directions. We shall need to return to this idea of internal energy when we discuss heat and molecular motion in Chapter 20.

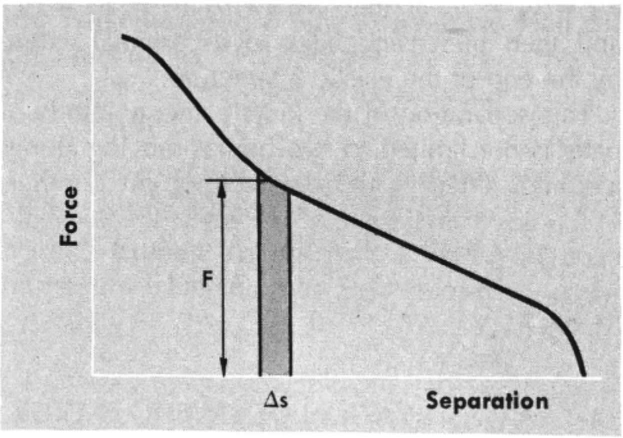
17–7 CONSERVATION OF KINETIC ENERGY IN ELASTIC INTERACTIONS

In the collision we studied in the last sections, the transfer of kinetic energy from one mass to the other ends up without loss. This result does not depend on either the range *d* or the magnitude *F* of the interaction force between the two masses. In any interaction with any interaction range, *d*, and any constant force, *F*, we find the same total kinetic energy at the beginning and the end of any collision. The result is more general than we might have anticipated. How general is it?

We shall now show that the total kinetic energy at the end of a collision is the same as at the beginning whenever the force of interaction depends only on the distance of separation of the two masses.* The equal and opposite forces *F* on the two masses may be any function of the separation of the two masses as long as *F* is zero beyond some definite range, so that we can define a complete interaction. Such an interaction starts with the masses farther apart than this range and ends when they are again separated by more than this range [Fig. 17 – 12 (a)].



17–12(a) The graph of a force which is a function of the separation of two bodies. The force is zero beyond the range of interaction.



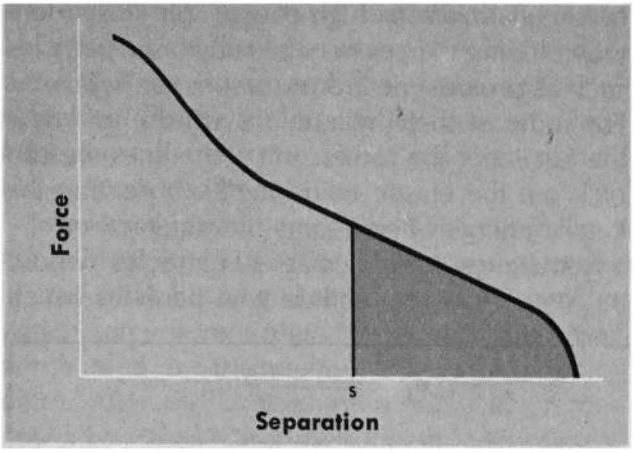
17–12(b) During an interaction, when the separation changes by Δs, the kinetic energy E_K changes by FΔs – the shaded area under the curve.

We can now apply the result $\Delta E_K(\text{total}) = F\Delta s$ to any force of interaction which depends on the separation alone—for example, the force illustrated in Fig. 17 – 12 (a). Let us assume that this is a repulsive force. Consider for a moment what happens to the total kinetic energy when the separation of the two bodies decreases by the Δs shown in Fig. 17 – 12 (b). The force is almost constant; and $F\Delta s$, represented by the shaded area, gives the decrease of the total kinetic energy. On the other hand, for this same separation, if the two bodies move apart by Δs, their total kinetic energy increases by the same amount, $F\Delta s$. We see then (Fig. 17 – 13) that as two interacting masses move closer together from large separation to separation *s* they lose all the kinetic energy represented by the area under the *F* versus *s* curve. Then, in going apart again, they gain back all the kinetic energy represented by the same area. At the end of the interaction they have the same kinetic energy as at the beginning, provided only that the force as a function of separation is the same when the masses go apart as it was when they came together. We have now extended our theorem of the conservation of kinetic energy over a completed collision to apply to any interaction in which the force depends only on the separation. An interaction of this kind is called an elastic interaction or an *elastic collision.*

The interactions we see about us, such as collisions between billiard balls, or between a baseball and a bat, are never perfectly elastic, but many are very nearly so. As we saw in Section 17 – 5, when a moving billiard ball strikes one at rest, practically no kinetic energy is lost. When we come to the physics of atoms and of elementary particles such as the electron, we shall find collisions which are even more nearly elastic. When one electron hits another one, no loss of

---

*This merely means that the force is the same on the way out as on the way in to the collision. This excludes forces like friction which reverse on the way out—so that they are always "against the motion."



17–13 If two bodies approach each other from a long distance to a separation s, they lose a total amount of kinetic energy equal to the shaded area under the curve. When they move apart again, they regain the same amount of kinetic energy provided only that the force as a function of separation is the same on the way out as on the way in.

kinetic energy is normally detectable, although in collisions with very rapidly moving electrons a noticeable amount of energy may be radiated away.

17–8 KINETIC ENERGY AND MOMENTUM

When two bodies interact elastically, their kinetic energies E_{K_1} and E_{K_2} change by equal and opposite amounts over the whole interaction; that is:

$$-\Delta E_{K_1} = \Delta E_{K_2}$$

This must be true because there is no change in the total kinetic energy $E_{K_1} + E_{K_2}$. Also, as we learned in Chapter 16, the changes in momenta of the two bodies are equal and opposite:

$$-\Delta\vec{\mathbf{p}}_1 = \Delta\vec{\mathbf{p}}_2$$

This vector relation is always true for two bodies forming an isolated system; in particular, therefore, it is true for elastic collisions.

The similarity between the equation relating the changes in momentum and the equation relating the changes in kinetic energy goes further. A change in momentum is given by the impulse $\vec{\mathbf{F}}\,\Delta t$, which is the measure of the transfer of momentum from one body to another, just as the transfer of energy is given by the work $F_x\Delta x$.

We can learn a great deal about the final motion of two interacting masses from the fact that neither momentum nor kinetic energy is lost in these two transfers. The two equations contain a large part of the information needed to determine the final velocities $\vec{\mathbf{v}}_1'$ and $\vec{\mathbf{v}}_2'$. In particular, if the motion takes place along a straight line, we find that

$$v_1' = \frac{m_1 - m_2}{m_1 + m_2} v_1$$

and that

$$v_2' = \frac{2m_1}{m_1 + m_2} v_1$$

For details of finding v_1' and v_2' for head-on collisions see the box on page 312. If the collision is not head on, we need some more information to tell us how far off the center it is and therefore in what direction the masses will go after the collision.

In making Fig. 17 – 8, we used a constant F within the range of interaction d; and we chose $m_1 = 3m_2$. Then from Newton's law we found that $v_1' = \frac{1}{2}v_1$ while $v_2' = \frac{3}{2}v_1$, as you can read off the figure. If you calculate v_1' and v_2' from the equations in the last paragraph based on the conservation of momentum and of kinetic energy, you will get the same results.

The results of using either Newton's law or the conservation laws must always agree. But it is not always so easy to apply Newton's law directly. If F is a complicated function of separation, finding the motion from Newton's law is a slow numerical procedure. Often, in fact, we do not know the details of F as a function of separation, even though experience may indicate that the total kinetic energy is the same before and after every collision. In such a situation, we cannot describe the motion completely, but the conservation laws still provide a large part of the answer.

For example, billiard-ball collisions are of this kind. The detailed forces between colliding balls are unknown to us; but observation shows that the kinetic energy of the balls is almost the same after they interact as it was before. Therefore, the forces must be functions only of the separation of the centers of the balls, and the forces are zero when the centers are apart by more than the diameter. In particular, then, we can use the last two equations with $m_1 = m_2$ to describe a head-on collision between a billiard ball in motion and one at rest. From the equations we then get $v_1' = 0$ and $v_2' = v_1$ or, in other words, if one billiard ball smacks another head on, the moving ball should stop and the ball at rest take off with the velocity the other used to have. That m_1 stops and m_2 goes on with the velocity v_1 is just what we observe, as Fig. 16 – 8 shows.

The same kind of analysis is often of importance in modern high-energy physics. Here the colliding masses may be subatomic particles such as protons, neutrons, mesons, or hyperons. For some of their interactions we do not know the details of the forces. But sometimes we can pick out the elastic collisions by observing the kinetic energies before and after interaction.

Sometimes visible masses (particles whose tracks we can see) collide with particles which

On the Derivation of the Formulas for v_1' *and* v_2'

For a head-on collision, in which the motions are along a straight line (as in Section 17 – 5), we can determine the final velocities v_1' and v_2' from the equations

$$-\Delta E_{K_1} = \Delta E_{K_2} \text{ and } -\Delta p_1 = \Delta p_2$$

(For vectors along a line, we drop the arrow and use the + or − sign to indicate direction, as in Chapter 11.) Writing ΔE_K as the difference between final and initial kinetic energies, we have

$$-(\tfrac{1}{2} m_1 v_1'^2 - \tfrac{1}{2} m_1 v_1^2) = (\tfrac{1}{2} m_2 v_2'^2 - 0)$$

We can factor this equation to give

$$-(m_1 v_1' - m_1 v_1)\left(\frac{v_1' + v_1}{2}\right) = (m_2 v_2' - 0)\left(\frac{v_2' + 0}{2}\right)$$

We see that each side of this energy equation is the product of the momentum change of one mass times the average of the initial and final velocities of that mass. Because we know that the momentum changes are equal and opposite, we can cancel $-\Delta p_1 = -(m_1 v_1' - m_1 v_1)$ on the left with $\Delta p_2 = (m_2 v_2' - 0)$ on the right. We therefore obtain the additional information that

$$v_1' + v_1 = v_2'$$

On substituting v_2' from this equation into the equation of the conservation of momentum

$$-(m_1 v_1' - m_1 v_1) = m_2 v_2'$$

we find

$$v_1' = \frac{(m_1 - m_2)}{m_1 + m_2} v_1$$

The energy and momentum relations thus lead to a specific prediction for the final velocity of m_1 when its initial velocity is given. Furthermore, by putting this value of v_1' into $v_2' = v_1' + v_1$, we get

$$v_2' = \frac{2m_1}{m_1 + m_2} v_1$$

which tells us the final velocity of the second mass.

Actually, in getting this answer for the final velocities we have put in quite a bit of information. Most of this information was introduced when we assumed that the collision was head on so that the masses move only along the x axis. If we had allowed one mass to approach the other off center (Fig. 17 – 6), the problem would be more complicated to handle, and we would have to put in the information representing the distance off center in order to get the answer. We also specified that m_2 is initially at rest. These equations do not apply if m_2 is in motion at the start of the interaction. The more general equations that apply then are, however, easy to derive by the same method.

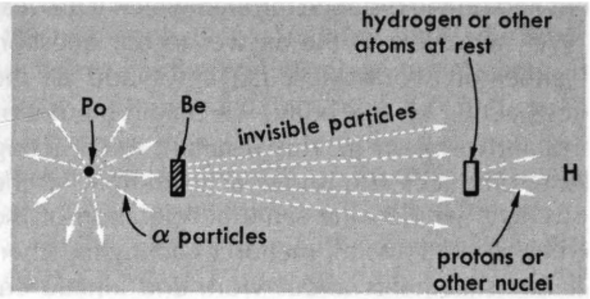
Furthermore, we should note that we made one tacit assumption in finding v_1' and v_2'. We assumed that there *was* a change in momentum in the interaction. It is possible even in our head-on collision to get another answer: if $v_1' = v_1$ and $v_2' = 0$, the momentum is unchanged and no energy transfers from one body to the other over the complete interaction. We shall not worry about this possible answer too much, however, because m_1 must pass right through m_2 in this case.

are invisible to us. If we assume that kinetic energy and momentum are conserved, and if we know the mass and the initial and final velocities of the visible particles, we can determine the mass and the initial and final velocities of the invisible particles.

In much this way James Chadwick discovered the neutron in 1932. In his experiments protons at rest were hit by unknown invisible particles which came from a region where beryllium was bombarded by the alpha particles from polonium. After the collisions the tracks of protons could be "seen" and their kinetic energy measured. After examining a large number of collisions, Chadwick concluded that one kind of invisible particle of a mass almost equal to the proton would explain all the observations. Furthermore, when the invisible particles hit atoms, the energy transferred was just what was expected. The "invisible" particle that Chadwick thus discovered is called the neutron. It is one of the building blocks out of which atoms are made; but its existence was only suspected before Chadwick's work. One of Chadwick's experiments is described in more detail in the box.

Discovery of the Neutron

Here is an outline of one of Chadwick's experiments leading to the discovery of the neutron. A sample of beryllium was bombarded by alpha particles (helium nuclei) from radioactive polonium, and the beryllium, in turn, was found to be giving off invisible, unknown particles (Fig. 17 – 14).



17–14 Diagram of Chadwick's apparatus.

These invisible particles then struck either hydrogen or nitrogen atoms at rest. As a result of the collisions, protons or nitrogen nuclei were knocked out, and Chadwick measured their velocities.

Suppose we select head-on collisions by looking in the region H (see Fig. 17 – 14), and we assume that the collisions are elastic. Then, calling m the mass of the invisible particle, and v its velocity, and calling m_p the mass of the protons and v_p' their velocity, we have

$$v_p' = \frac{2m}{m + m_p} v$$

In the same way, calling the mass of nitrogen m_N and its velocity v_N,

$$v_N' = \frac{2m}{m + m_N} v$$

In this equation we can replace m_N by $14m_p$ because a nitrogen atom has 14 times the mass of a hydrogen atom. After replacing m_N by $14m_p$, we divide the first equation by the second one. This eliminates the unknown velocity v of the invisible particle with the result that

$$\frac{v_p'}{v_N'} = \frac{m + 14m_p}{m + m_p}$$

In his experiments, Chadwick measured the velocities v_p' and v_N'. He found that the ratio $\frac{v_p'}{v_N'}$ was about 7.5. Therefore

$$\frac{m + 14m_p}{m + m_p} \approx 7.5$$

or

$$m \approx 1.00 m_p$$

Chadwick repeated the experiment with other substances in place of hydrogen or nitrogen and found again that an invisible particle with about the mass of a proton fitted in with his measurements. He did a number of other experiments. All were consistent, and one of them determined the mass to within less than 1 percent. He had proved that the neutron exists.

17–9 WORK AND KINETIC ENERGY WHEN MORE THAN ONE FORCE ACTS

In much of this chapter we have studied the gains and losses of the kinetic energy of a body when it is acted on by a single external force. Even when we dealt with the interaction of two bodies, each body was acted on by a single force. If we now consider more than two bodies, each body will be acted on by the forces of interaction with every other body. How then do we find the changes in kinetic energy of the various bodies?

Let us consider one of the bodies. How does the work which transfers kinetic energy in or out of the body depend on all the forces acting on it and on the distance through which it moves?

We can arrive at the answer to our question in either of two ways. First, if we add all the forces acting on the body, we obtain the net or resultant force on it. This resultant force acting alone produces the same acceleration of the body as a whole—the same acceleration of the center of mass—as all the forces acting together. All our statements about work and kinetic energy are true if the force we use is the *net* force. The kinetic energy we get this way is the energy of motion of the body as a whole, excluding the energy of internal motions like vibration or rotation of the parts about the center of mass.

Second, we can compute the work done by any one of the forces acting during a specified part of the motion, and add together the amount of work done by each. The sum will be just equal to the work done by the net, or resultant, force. To see how this result comes about, suppose that the motion of the body is along the x axis. Then the work done by each of the forces F_a, F_b . . . etc., acting on the body is

$$W_a = F_{ax}x, \qquad W_b = F_{bx}x. \; . \; . \; .$$

In each case it is the component of the force in the direction of the motion times the distance x which the body moves. The total work is then the sum

$$\begin{aligned} W &= W_a + W_b + \ldots \\ &= (F_{ax} + F_{bx} + \ldots)\, x = F_x x \end{aligned}$$

where F_x is the component of the net force along the direction of motion.

In physical language each force represents the push or pull of some outside agency on the moving body. If the force is in the direction of motion, the work done by this force represents the transfer of energy from the outside agency to the moving body. If the force is opposite to the direction of motion, the transfer is from the moving body to the outside agency. The net result of the whole process will depend on whether the resultant force is in the direction of motion or opposite to it. In the first case more energy is transferred to the moving body than is taken from it. In the second case more energy is taken out of the moving body than is supplied to it. In this way we can generalize the results of this chapter to apply to as many bodies or as many external forces as are present in a complex situation.

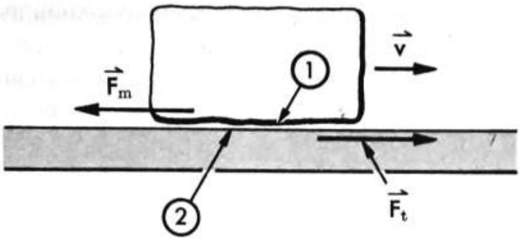
17–10 LOSS OF KINETIC ENERGY IN A FRICTIONAL INTERACTION

If a frictional force acts on a body—a force that does not depend just on the separation between two bodies—energy is apparently turned into some other form. The kinetic energy of the overall motion of the interacting bodies decreases and finally becomes zero.

Consider the motion of a mass sliding on a tabletop. The frictional force exerted on the mass by the tabletop slows the mass down. The work done is the product of the force and the distance the mass moves. Because the force is opposite to the motion, it transfers energy out of the mass and hence the kinetic energy of the mass decreases by the amount of this transfer.

To what is the energy transferred? There is an equal and opposite force exerted by the mass, acting on the table, but the table hardly moves under the influence of this force. Consequently this force on the table does practically no work, and the kinetic energy of the table does not increase appreciably. The kinetic energy of the mass is taken out, but it is not put into the kinetic energy of the table (Fig. 17 – 15).

It is, however, common observation that the surfaces of the mass and of the table which have been in contact become warm. The energy taken out of the moving mass is associated with this increase in temperature. We can already



17–15 A mass is sliding along an ordinary table with speed v. A horizontal frictional force F_m acts in region 1 on the lower surface of the mass, taking kinetic energy out of it. The loss of E_K is measured by the work done as the mass moves, slowing down as a result of this frictional force. A frictional force F_t on the table is exerted in region 2 on the surface of the table; it does no work, since the table does not move. The table gains no E_K. Where has this energy gone? Apparently it accumulates in the region of contact around 1 and 2, where we observe an increase in temperature.

get some idea of the connection between this temperature increase and the missing kinetic energy when we recall that the temperature of a gas is proportional to the average mv^2 of its molecules. This means that warming a body, raising its temperature, increases the kinetic energy of its molecular motions. The energy which does not appear as kinetic energy of the motion of the table as a whole may thus be found in the increased energy of motions of the molecules in the warmed regions of the table and the mass.

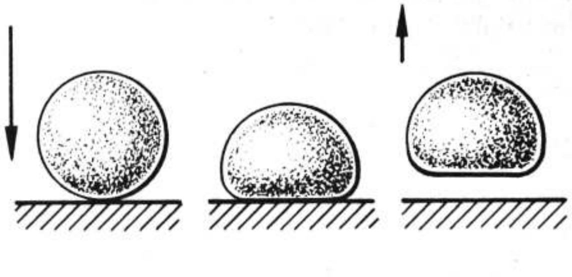
Let us consider another example of a frictional interaction. When we drop a ball of putty, the interaction between the putty and the floor begins as soon as the ball and the floor come into contact. Forces then begin to act on the putty, slowing it down. Also, the shape of the ball is changed; and, on the rebound, the interaction stops when the ball is closer to the floor than when the interaction began (Fig. 17–16). The force between ball and floor at this position is now zero, whereas when the ball was moving down at this position, the force was not zero. As a result, the total kinetic energy is less after collision than before and the ball of putty bounces back, moving very slowly. The putty ball seems to lose kinetic energy permanently, but we find the putty warmer after collision with the floor than it was before. This interaction is much like that between the sliding mass and the table. In this example, however, the friction may all occur inside the putty ball when one part of the ball moves with respect to another.

A collision in which the forces are smaller when the bodies separate than when the bodies come together is called an *inelastic* collision. After inelastic collisions, a temperature rise is usually observed. In Chapter 20 we shall discuss how this increase in temperature is related to energy in another form. For the present we shall keep our attention on situations in which friction and heating effects can be neglected.

17–11 CONCLUSION

What are the main things that we have found out in this chapter? In the first place, when a force is applied to a moving body, the transfer of energy to the body is given by the work: that is, by the distance the body moves times the component of the force in the direction of the motion. If the force is opposite to the motion, energy is taken from the body. If the force is in the direction of motion, energy is fed into the body. When more than one force acts on a body the net force changes the motion of the body; and correspondingly the total work measures the change in its kinetic energy.

When two bodies interact and the force of interaction depends only on the distance between the bodies, the total kinetic energy after a completed interaction is the same as the total kinetic energy before the interaction. The collision of the two bodies is elastic.



17–16 An inelastic collision. The interaction begins when the ball of putty comes in contact with the floor. The ball is permanently deformed by the interaction and, on the rebound, the interaction ceases before the ball reaches the separation at which the interaction started. As a result, the speed on the rebound is less than the speed at which the ball came in.

We can make the same kind of statement when many bodies interact. If they all start so far apart that the forces of interaction are zero and all end up far apart again, the kinetic energy at the beginning and at the end is the same. Some bodies may have gained kinetic energy, others may have lost it, but the total comes back to what it was. This statement about the equality of the total kinetic energy at the beginning and at the end of a completed interaction will be proved in general in the next chapter. Here we should emphasize that the statement is correct only if the interaction forces are functions of the separation of the bodies alone. In other words, the forces must be the same whether the bodies are approaching each other or moving apart.

Forces of interaction are not always functions of separations alone. Recall the examples of the mass rubbing on the table and the putty ball bouncing on the floor. They show that when the forces depend on other things the kinetic energy of bulk motion seems to change permanently. But when this kinetic energy disappears a temperature rise is usually observed. Later we shall find how this temperature increase indicates that the kinetic energy has been transferred to another form.

Even in the course of an elastic collision, the total kinetic energy of the bodies may appear to be lost temporarily. We have found that this lost kinetic energy can eventually be regained. As long as the forces of interaction depend only on the relative positions of the bodies, the "lost" kinetic energy is actually stored in the system. This stored energy we call *potential energy*. It is the subject of the next chapter.

QUESTIONS

1. Coal is burned in a city electric power plant to make steam to run a steam turbine that drives an electric generator. The city water department uses this electric supply to run an electric motor to pump water from a well into a standpipe on a hilltop. List, in order, all the energy changes that take place.

2. A cable pulls a car up a mountain with a force of 4×10^3 newtons at a velocity of 5 m/sec. It takes the car 5 minutes to reach the top.

   (a) How much work is done in getting the car to the top of the mountain?

   (b) How much work would be done to get the car up the mountain if it traveled 2.5 m/sec?

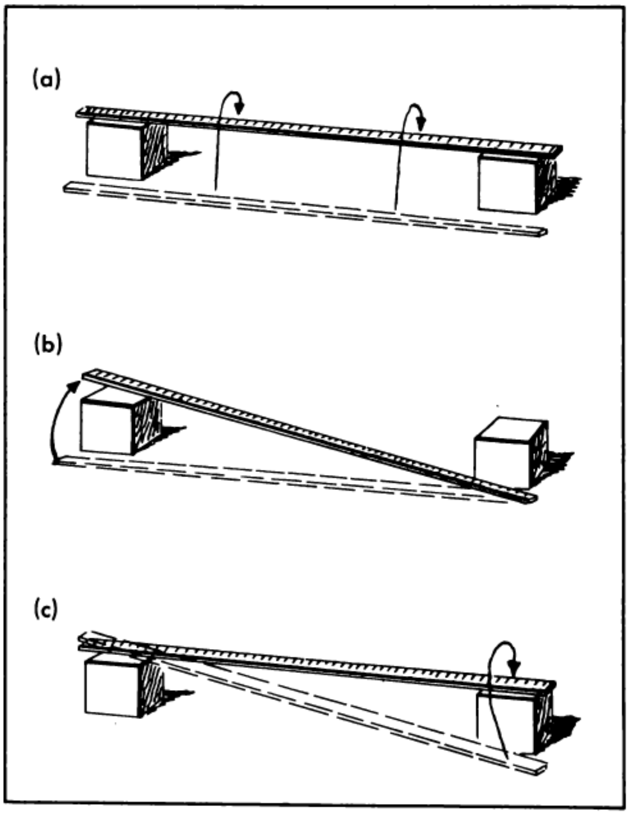
3. A boat engine delivers energy at a rate of 3000 joules/sec (power of 3000 watts), driving the boat at a constant speed of 9 km/hr. What is the force resisting the motion of the boat?

4.* Is it possible to exert a force and yet transfer no energy?

5. A meter stick, mass 0.20 kg, is lying on a table near two blocks 10 cm high (Fig. 17 – 17).

   (a) If you lift the stick, holding it horizontal, and put it on the blocks, how much work have you done?

   (b) If you lift one end and set it on one block, and then lift the other end, setting it on the other block, how much work have you done in moving the stick?



17–17 **For Question 5.**

6.* What is the kinetic energy of a 1-kg hammer moving at 20 m/sec?

7.* How does the kinetic energy of a car change when its speed is doubled?

8.* Two bodies of unequal mass each have the same kinetic energy and are moving in the same direction. If the same retarding force is applied to each, how will the stopping distances of the bodies compare?

9. A force of 10.0 newtons acts on a 2.00-kg roller skate initially at rest on a frictionless table. The skate travels 3.00 meters while the force acts.
   (a) How much work is done?
   (b) How much energy is transferred to the skate?
   (c) What is the final speed of the skate?

10. Compare the kinetic energies of two objects, *A* and *B*, identical in every respect except one. Assume that the single difference is:
   (a) Object *A* has twice the velocity of *B*.
   (b) Object *A* moves north, *B* south.
   (c) Object *A* moves in a circle, *B* in a straight line.
   (d) Object *A* is a projectile falling vertically downward; *B* is a projectile moving vertically upward at the same speed.
   (e) Object *A* consists of two separate pieces attached by a light string, each equal in mass to the mass of *B*.

11. A force of 30 newtons accelerates a 2.0-kg object from rest for a distance of 3.0 meters along a level, frictionless surface; the force then changes to 15 newtons and acts for an additional 2.0 meters.
   (a) What is the final kinetic energy of the object?
   (b) How fast is it moving?

12. Estimate your kinetic energy (in joules) when you ride your bicycle on the road.

13. A 2.0-kg stone whirls around on the end of a 0.50-m string with a frequency of 2.0 revolutions per second.
   (a) What is its kinetic energy?
   (b) What is the centripetal force on it?
   (c) How much work is done by the centripetal force in one revolution?

14.* In Fig. 17 – 8, do the two bodies repel or attract each other?

15. A 10.0-kg mass moves 2.00 meters against a retarding force that increases linearly by 4.00 newtons for every 3.00 meters the mass moves (see Fig. 17 – 18). If the force is zero at the beginning, how much kinetic energy is lost?

16.* When do we call a collision *elastic*?

17.* When a putty ball is thrown at a wall and sticks to it, does the force between the wall and the putty ball depend on the separation alone?

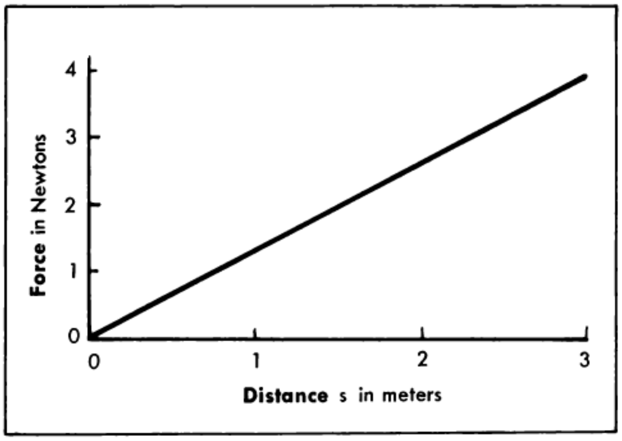
18.* A 3-kg ball with a speed of 6 m/sec strikes another 3-kg ball moving 6 m/sec in the opposite direction. After collision each ball has a velocity equal in magnitude but opposite in direction to its initial velocity. Is this an elastic collision?

19.* A rapidly moving mass m_1 strikes a mass m_2 at rest in a head-on elastic collision. How does m_1 compare with m_2 if m_1 comes back along its path at a very low speed?

20. Suppose two bodies collide. When they approach from 10 to 5 cm the repulsive force between them is 3 newtons, and when they go apart from 5 to 10 cm the repulsive force is 1 newton. What will be the change of the total kinetic energy after the cycle of approaching and retreating from 10 to 5 and from 5 to 10 cm?

21. A projectile from an antitank gun (which does not contain an explosive charge) has a mass of about 10 kg and a speed of 10^3 m/sec. A freight car being moved around in a switchyard has a mass of about 10^4 kg and a speed of 1 m/sec. What are their momenta? Their kinetic energies? Why does the projectile do much more damage than the freight car when it hits something?

22. Two 3.0-kg bodies interact. At a given moment the first body is moving to the right at 0.50 m/sec and the second body is moving to the right at 0.30 m/sec, and so at that instant the speed of approach of one body as measured from the other is 0.20 m/sec. If the force of interaction is repulsive and equal to 0.10 newton, at what rate is the total kinetic energy decreasing at that time?



17–18 For Question 15.

23. A 1.5-kg body is at rest. It is "hit" head on by a body of mass 0.50 kg moving with a speed of 0.20 m/sec. The interaction force depends only on the separation of the two bodies.

(a) What is the final velocity of each body?

(b) In what direction does each move after the interaction?

24. An object of mass m_1 and kinetic energy E_K collides head on with an object of mass m_2 at rest. Assuming the force of interaction depends only on the separation, calculate the kinetic energy transferred if

(a) $m_2 = .01\ m_1$

(b) $m_2 = m_1$

(c) $m_2 = 100\ m_1$

What is the ratio of m_1 and m_2 when the transfer of kinetic energy is

(d) a maximum?

(e) very small?

(*Note:* To see this relationship more clearly, you may wish to find the energy transferred for additional values of m_2, and make a graph.)

25. A 5.0-kg body is at rest. A 10-kg body approaches it with a velocity of 0.20 m/sec. The interaction force is zero when the separation is greater than 0.10 meter and is 4.0 newtons when the separation is less than this distance. (Note that some of the answers in this problem are easier to get if you use the conservation of momentum.)

(a) What is the kinetic energy of the masses before the interaction?

(b) What will be the kinetic energy of each mass after the interaction is complete?

(c) What will be the kinetic energy of each mass when the separation is at a minimum? Recall that the velocities at minimum separation are equal.

(d) What is the minimum separation? [Your answer to part (c) tells you the net loss of kinetic energy of the two masses at minimum separation.]

26. An antiaircraft shell has a kinetic energy E_K and a momentum $\vec{\mathbf{p}}$. Just then it explodes. What can you say about:

(a) the momentum of the pieces?

(b) the kinetic energy of the pieces?

27. In one of the experiments that led to the determination of the mass of the neutron, Chadwick measured the velocity of protons that had been hit head on by neutrons. The velocity of the protons was 3.3×10^7 m/sec.

(a) What was the velocity of the neutrons before and after collision with the protons?

(b) Chadwick also measured the velocity of nitrogen atoms hit head on by the neutrons. What was it?

(c) What was the velocity of the neutrons after each kind of collision?

28. An empty freight car A of mass 2.5×10^4 kg (25 tons) coasts along a horizontal track at 2.0 m/sec until it couples to a stationary car B of mass 5.0×10^4 kg. There is little friction, and the brakes are off.

(a) What is the initial momentum of car A?

(b) What is the speed of the two cars as they move along the track after the interaction?

(c) What is the total kinetic energy before and after the impact? Is this an example of an elastic collision?

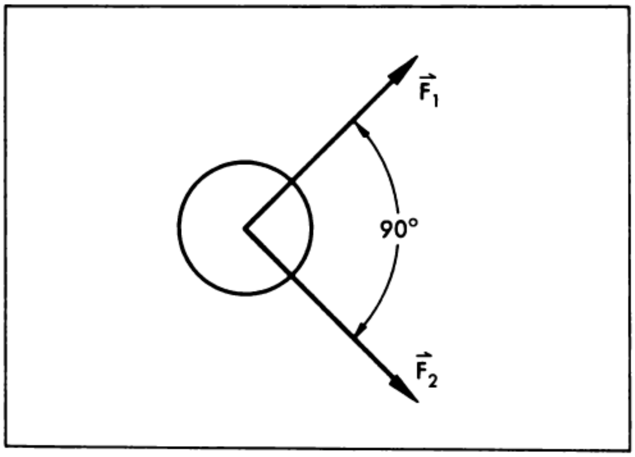
29. A Dry Ice puck is acted upon by two forces $\vec{\mathbf{F}}_1$ and $\vec{\mathbf{F}}_2$ as shown in Fig. 17 – 19. Assume that in each case the body starts from rest.

How much work is done on the body as it moves 2 meters if:

(a) $F_1 = 10$ newtons and $F_2 = 0$?

(b) $F_1 = 0$ and $F_2 = 10$ newtons?

(c) $F_1 = 10$ newtons and $F_2 = 10$ newtons?



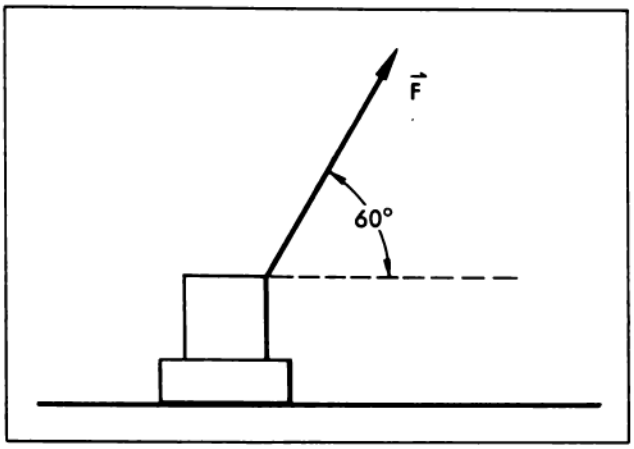
17–19 For Question 29.

30. A Dry Ice puck has a mass of 5 kg. It is subjected to a constant force $\vec{\mathbf{F}}$ of 50 newtons at an angle of 60° above the horizontal for 2 seconds (Fig. 17 – 20).

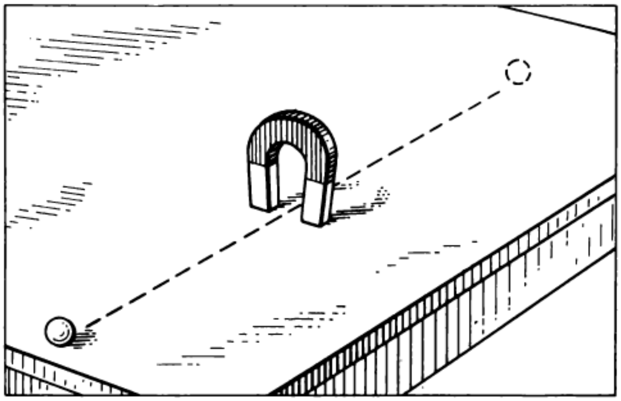
(a) If the puck starts from rest, what is its change in momentum in the first two seconds the force is applied?

(b) How much work was done in these two seconds? (*Note:* $E_K = \frac{1}{2}mv^2 = \frac{(mv)^2}{2m} = \frac{p^2}{2m}$.)

(c) What would happen if the force were doubled?



17–20 **For Question 30.**



17–21 **For Question 33.**

31. A 100-gm light bulb dropped from a high tower reaches a velocity of 20 m/sec after falling 100m. About how much energy has been transferred to the air?

32. Suppose you wish to determine the force your bicycle brakes are exerting on the bicycle when you brake hard without skidding. Assume the brakes grip the rim of the wheel. At your disposal is a tape to measure length and a road of known constant slope. Also, you know the total weight of the bicycle and yourself. You do not have a watch. How would you go about it?

33. A horseshoe magnet of mass *m* stands on end on a frictionless table. A steel ball bearing of mass *m* is rolled toward the magnet from far away with velocity *v* and goes through the magnet and far beyond. Assume the force of attraction *F* changes with distance and is the same in front and behind the magnet. (See Fig. 17 – 21.)
    (a) What is the final velocity of the ball?
    (b) What is the final velocity of the magnet?

34. After the experiment described in Question 33 has been done with the starting velocity *v*, the magnet and ball are returned to their original positions and the ball is sent in with a velocity v' such that $v' > v$. Will there be any difference in the motion of the magnet?

35. What would be the kinetic energy of a 2.0-kg ball moving in a circle of 1.0-m radius 300 times per minute? What would the kinetic energy be if the ball were split into four equal parts, each 90° apart on the circle, with the same motion as before? Would the kinetic energy change if the 2.0-kg ball were formed into a continuous ring 2.0 m in diameter turning at the same rate?

18

Potential Energy

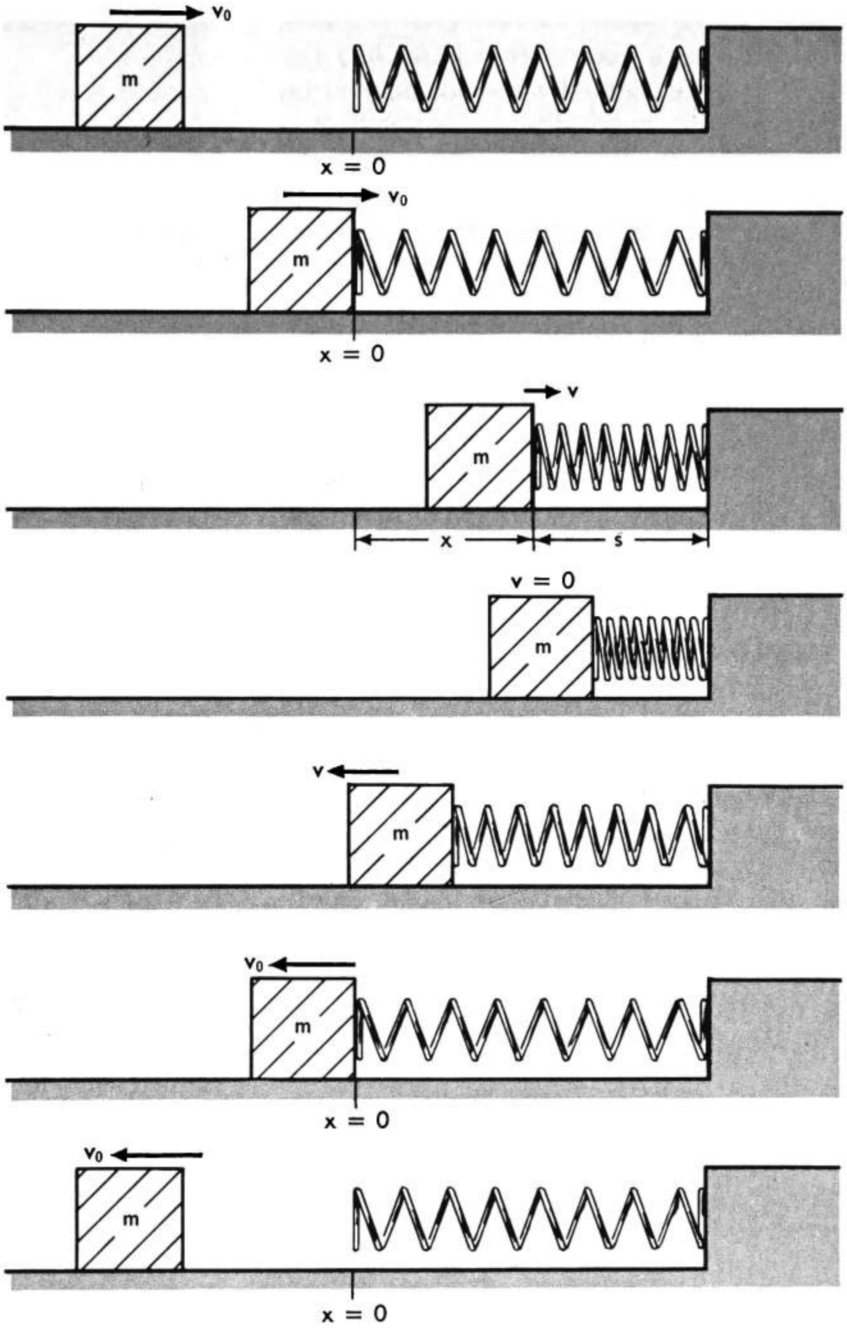
IN THE LAST chapter we focused our attention on the transfer of kinetic energy from one moving body to another. We found that for completed collisions within an isolated system the loss of kinetic energy of one body must equal the gain of kinetic energy of the other *as long as the force of interaction depends solely on their separation*. Then with the help of the law of conservation of momentum we were able to calculate the final velocities of two bodies colliding head on in terms of their masses and their velocities prior to the interaction.

In this application, the law of conservation of momentum, $-\Delta\vec{\mathbf{p}}_1 = \Delta\vec{\mathbf{p}}_2$, and the law of conservation of kinetic energy, $-\Delta E_{K_1} = \Delta E_{K_2}$, seem strikingly similar; but there is an important difference. The changes in momentum are equal and opposite in any time interval, and the momentum is therefore conserved instant by instant throughout the interaction. On the other hand, even in elastic collisions the total kinetic energy is not the same at all stages of the interaction. Only at the end of the interaction does it return to its initial value. During the collision, the total kinetic energy first decreases and then increases. At intermediate stages some of the kinetic energy has disappeared.

What happens to this lost kinetic energy? Because it all comes back, it must be stored somehow in the interacting system. We call this stored energy the *potential energy* of the system.

18–1 THE SPRING BUMPER

Here is a simple example of energy being stored. Consider a mass m sliding with constant velocity on a horizontal frictionless table (Fig. 18 – 1). The mass collides with a spring bumper attached to a large body so massive that it hardly moves. When the moving mass hits the spring, the spring is compressed. It exerts a force back on the moving mass, slowing it down. The kinetic energy of the moving body decreases until the speed is zero. At this point the kinetic energy of the moving body has disappeared, and the spring is compressed a maximum amount.



A mass *m* is approaching a spring bumper with speed v_0.

It hits the spring and starts to compress it.

When the spring is compressed a distance *x* to separation *s*, the speed of the mass has decreased to *v*. The mass has lost kinetic energy, which has been stored as potential energy by the compressed spring.

At maximum compression, the mass has come to rest. All its kinetic energy has disappeared.

As the spring is still shoving, the mass gains speed and kinetic energy.

The mass has returned to the place where it first hit the spring. It now has its original speed v_0, and its original kinetic energy. The interaction is completed.

The mass continues to move away with speed v_0 and its original kinetic energy.

18–1 A collision between a mass m and a spring attached to a mass so large that it does not move appreciably.

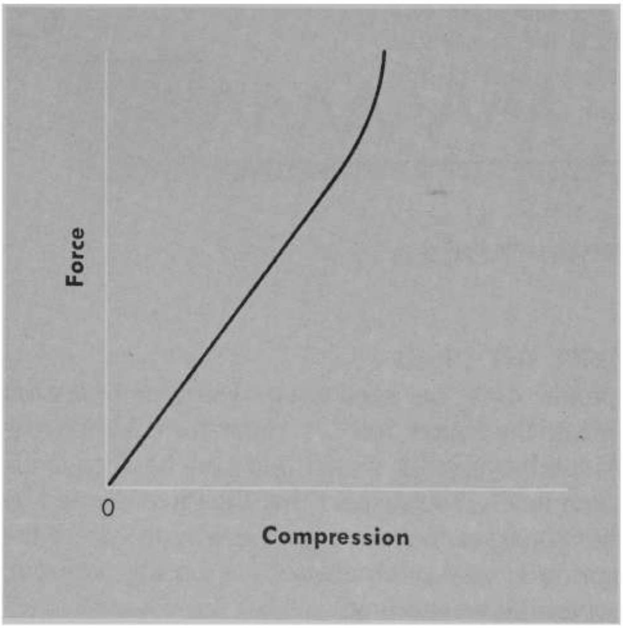
All the energy is stored as potential energy. After that the mass picks up speed in the opposite direction. Finally it leaves the spring with its original speed and kinetic energy. All the kinetic energy lost during the compression has been regained. At intermediate compressions, the energy was partly kinetic and partly potential.

In the last chapter we saw that kinetic energy will be completely regenerated if the force depends only on separation. Here compression plays the same role as separation. When the kinetic energy is stored and can be recovered completely, we suspect that the force exerted by the spring is the same on the way in, while the spring is being compressed, as on the way out, when it is expanding.

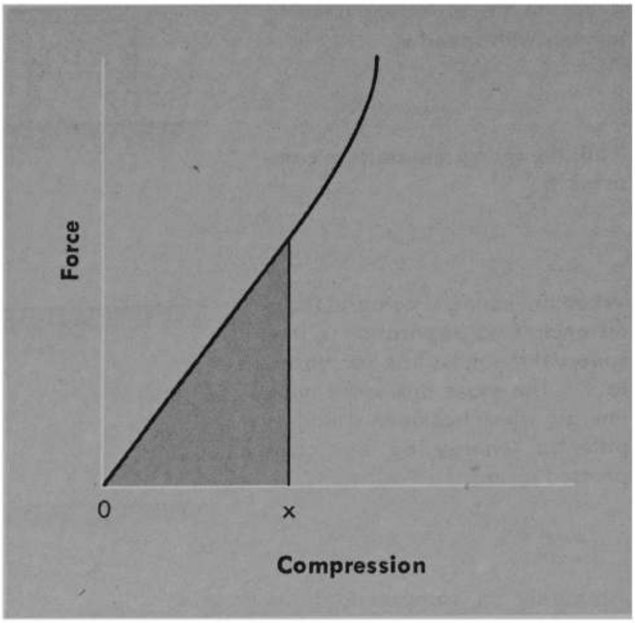
Measurements of the force as a function of compression confirm our suspicion. As we saw in Chapter 14, a typical curve for the restoring force exerted by a good spring looks like Fig. 18 – 2. The force exerted does not depend on the past history. It has the same value at a given compression if we have just pushed the spring in that far or if we have pushed it in farther and then let it expand. Furthermore, the force does not depend on the speed of the mass. Since the force is always the same at the same compression, we can represent force by a single curve, as we have done in Fig. 18 – 2.

On the other hand, for a poor spring—a copper coil, for instance—the history matters. The forces on the way in and on the way out differ. When such a coil is hit by a moving mass, the mass bounces away, moving more slowly than when it came in. A copper coil is inelastic like putty and warms up when it is moved in and out. Therefore, we do not get a single curve of force versus compression.

When a mass hits a spring, compressing it and losing kinetic energy, the transfer of energy from the kinetic energy of the mass into potential energy stored in the compressed spring is measured by the work. This work is represented by the area under the force-compression curve from zero (when the spring is uncompressed) to x (the distance through which the end of the spring has moved; Fig. 18 – 3). When we are dealing with a good spring, one with a single force-compression curve, this work is always the same when the same compression is reached. The *loss* of kinetic energy by the moving mass is therefore always the same. It does not matter what the original kinetic energy of the mass is. If m comes in with higher kinetic energy, it will have higher kinetic energy as it is passing x; but the change ΔE_K between zero and x is the same. This *loss* of kinetic energy, this work done in compressing the spring, is the potential energy stored in the spring.



18–2 A graph of the restoring force F exerted by a good spring, as a function of its compression.



18–3 The work done in compressing a spring is measured by the area under the F versus x curve from zero to any compression x. This represents the potential energy stored in the spring at the compression x.

It does not matter whether a moving mass compresses the spring or whether we compress it. If we compress the spring a distance x by hand, place the mass at its end, and let go, the spring will flip the mass off with kinetic energy equal to the work we did by hand. The potential energy of the spring is again given by the area under the force-compression curve, and it is a property of the compression of the spring without any reference to the moving mass.

For a spring with a given force-compression curve we can evaluate the potential energy as a function of the compression x by finding the areas under the curve between zero and various values of x. If we plot these areas against x, we obtain a graph of the potential energy U. Figure 18 – 4 is such a graph of U versus x for the spring with the force-compression curve of Fig. 18 – 2.

If the force-compression curve is simple enough, we can find a formula for the potential energy U as a function of x. In Fig. 18 – 5 we see that when the force is proportional to the compression, $F = kx$, the area under the curve is a triangle with base x and altitude kx. The potential energy is therefore

$$U = \tfrac{1}{2}(\text{altitude})(\text{base})$$
$$= \tfrac{1}{2}(kx)x = \tfrac{1}{2}kx^2$$

This formula gives the potential energy for a spring with a linear restoring force.

The graph of U versus x can be checked experimentally (whether it was obtained from a formula or from the areas under the force-compression curve). Compress the spring by different amounts x and let it go, each time accelerating a known mass. Measure the kinetic energy with which the mass ends up. Its value should be the same as U at compression x.

The kinetic energy of the mass when it leaves the spring is equal to the potential energy when the mass is at rest, because all the potential energy is turned into kinetic energy on the way out. At intermediate points the gain in kinetic energy and the loss of potential energy are equal, and the sum is constant: that is,

$$\tfrac{1}{2}mv^2 + U = E$$

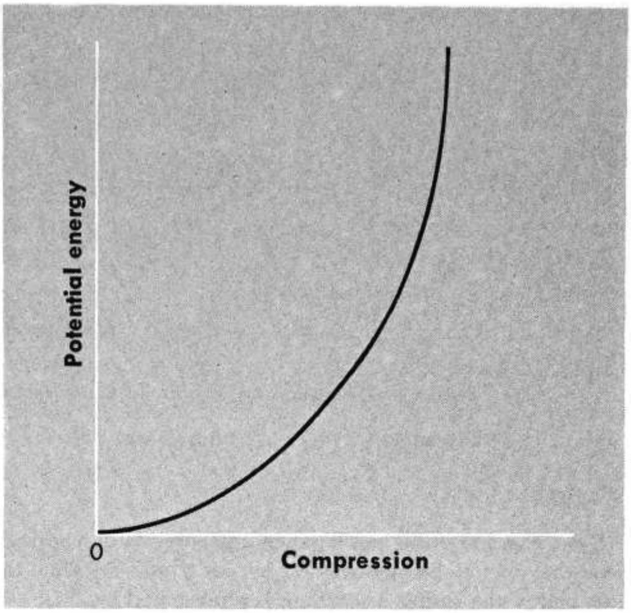
The constant E is the potential energy at maximum compression when $v = 0$. It is also equal to $\tfrac{1}{2}mv_0^2$, the kinetic energy when $U = 0$ and the mass is just leaving the spring. We call E the total energy of the spring and the mass.

When we fasten a mass m to the end of any spring and set the system in motion, it oscillates back and forth. The expression

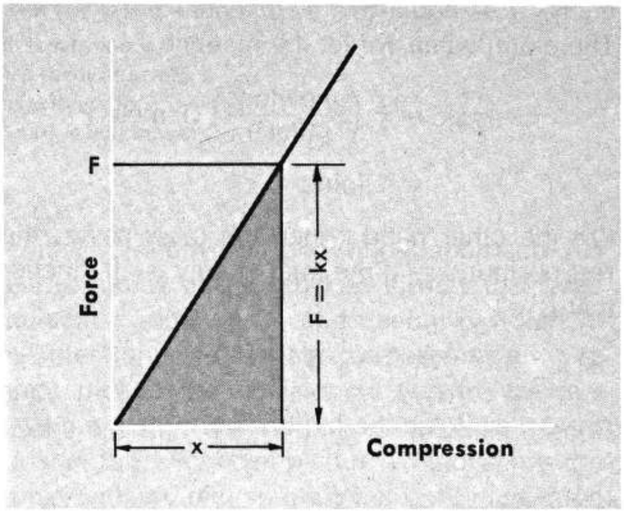
$$\tfrac{1}{2}mv^2 + U = E$$

enables us to calculate the speed v from the potential energy U at any stage of the motion (Fig. 18 – 6). For example, if the spring exerts a *linear* restoring force, we know that the mass oscillates in simple harmonic motion (Section 14 – 8). Also we know that $U = \tfrac{1}{2}kx^2$ for such a force. For simple harmonic motion, then, we see that

$$\tfrac{1}{2}mv^2 + \tfrac{1}{2}kx^2 = E$$

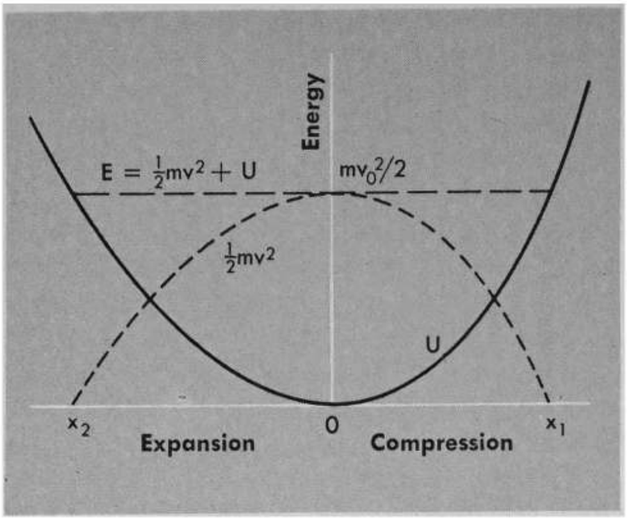


18–4 The graph of potential energy versus compression for the spring with the force-compression curve of Fig. 18–2.



18–5 The graph of F versus x for a spring, when F = kx. The area under the curve is the triangle with base x and height kx. The area, and thus the potential energy, is given by 1/2 kx².

For example, suppose we have a spring of force constant $k = 2$ newtons/meter. We attach a mass $m = 8$ kg to it and pull it out 1 meter from its rest position. Then we let the mass go. What will be its speed as it passes through the rest position? With what period will the mass oscillate back and forth? To answer the first question, we first notice that the mass is not in motion when we let it go. The kinetic energy $\tfrac{1}{2}mv^2$ is therefore zero, and the constant total



18-6 A graph of the potential energy curve U of a spring and the kinetic energy curve 1/2 mv² of a mass attached to the end of the spring. The sum E is represented by a horizontal line, indicating that the sum of the energies is constant. The mass has just enough energy to compress the spring to $x = x_1$ or expand it to $x = x_2$.

energy E is equal to the potential energy $\frac{1}{2}kx^2$. Therefore, when we let the mass go

$$E = \frac{1}{2}kx^2 = \frac{1}{2}\left(2\frac{\text{newtons}}{\text{meter}}\right)(1\ \text{meter})^2$$
$$= 1\ \text{joule}$$

On the other hand, when the mass passes the rest position $x = 0$ the total energy is all kinetic; that is

$$E = \frac{1}{2}mv^2 = \frac{1}{2}(8\ \text{kg})v^2 = 1\ \text{joule}$$

Consequently, $v^2 = \frac{1}{4}\left(\frac{\text{meter}}{\text{sec}}\right)^2$ and the speed is $\frac{1}{2}$ m/sec.

We can answer the second question—what is the period?—in two simple ways. First we know from Section 14-8 that simple harmonic motion is the motion of a component of steady circular motion. If the maximum displacement in the simple harmonic motion is 1 meter, the matching circular motion goes around a circle of 1 meter radius and, hence, 2π meters circumference. Also, if the maximum speed in simple harmonic motion is $\frac{1}{2}$ m/sec, the steady speed of the matching circular motion is also $\frac{1}{2}$ m/sec. Consequently, the period is

$$\frac{2\pi\ \text{meters}}{\frac{1}{2}\ \text{m/sec}} = 4\pi\ \text{sec}$$

We can check this result by the second method. In Section 14-8 we learned that the period is

$$T = 2\pi\sqrt{\frac{m}{k}}$$

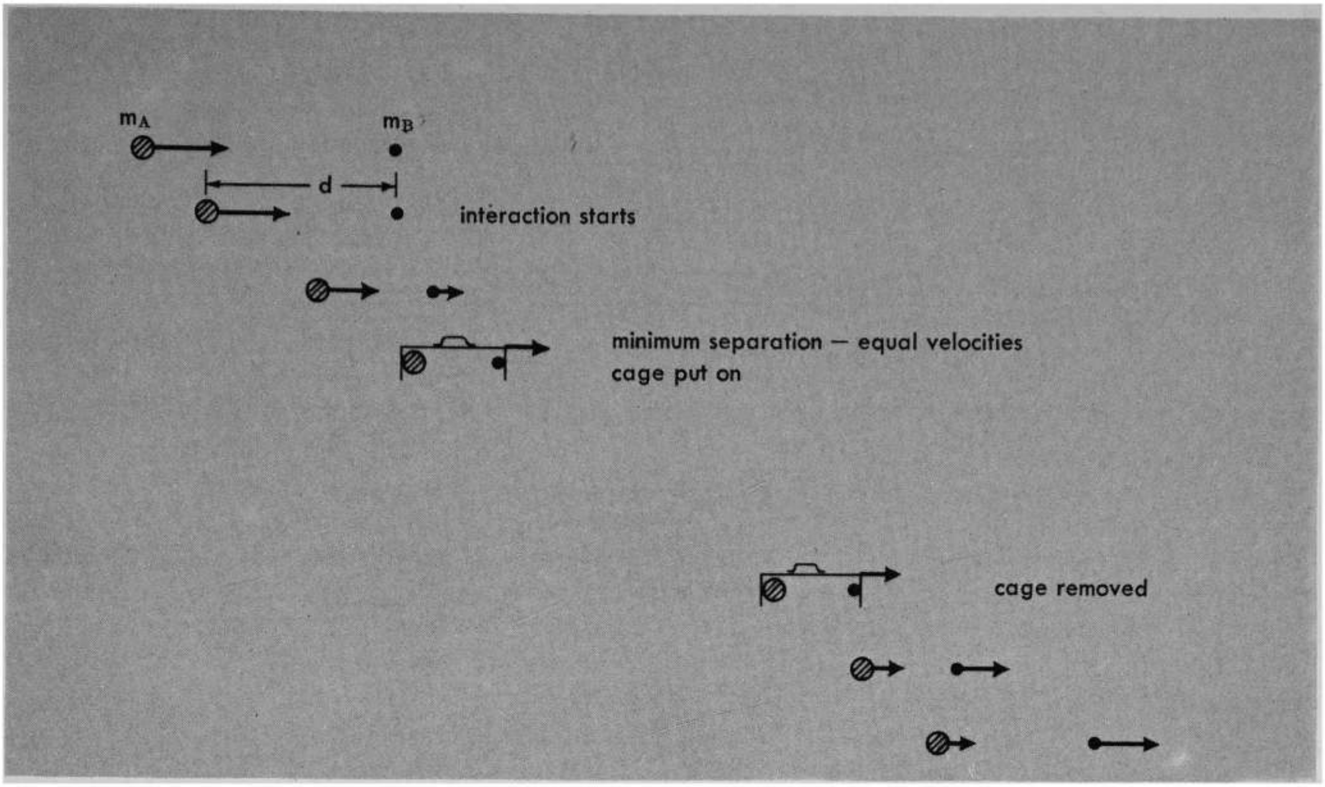
Here this is $T = 2\pi\sqrt{\dfrac{8\ \text{kg}}{2\ \text{nt/m}}} = 4\pi$ sec

which agrees with the period as computed by the first method.

18-2 POTENTIAL ENERGY OF TWO INTERACTING BODIES

Now consider one mass colliding head on with another, as in Chapter 17. Suppose mass A is projected toward mass B, which is at rest (Fig. 18-7). As we learned in Section 17-5, when A reaches the range of interaction d, it begins to slow down and lose kinetic energy, while B starts moving faster and faster; but B does not gain as much kinetic energy as A loses. Kinetic energy is disappearing. When A and B are closest together they are both moving with the same velocity, and the total kinetic energy is at a minimum. Suppose that at just that instant a light cage (of negligible mass) is dropped over A and B to prevent them from separating. Then the whole system—A and B and their cage—continues with unchanging velocity and constant kinetic energy. It will continue to move with this minimum kinetic energy until we remove the cage. If we look at kinetic energy alone, the system seems to have less energy than it had originally. However, if we remove the cage so that A and B can push apart, we get back to the original kinetic energy, provided that the interaction forces depend on the separation alone. Here we are considering only those interactions which provide the same forces at the same separation, whether A and B are approaching or moving apart. Such forces make the collision elastic.

While the cage is on, we have less kinetic energy than we had originally, and we say that the missing kinetic energy is stored as potential energy. We can keep it stored as long as we like. And whenever we allow the two interacting masses to move apart again, the kinetic energy will increase by just the amount that we



18–7 An interaction between two masses, which begins when their separation is d. At their closest approach, when their velocities are the same, we enclose them in a very light cage. They will continue to move along together at the same velocity, with minimum kinetic energy. We can remove the cage at any time; the masses will move apart again and regain the original kinetic energy.

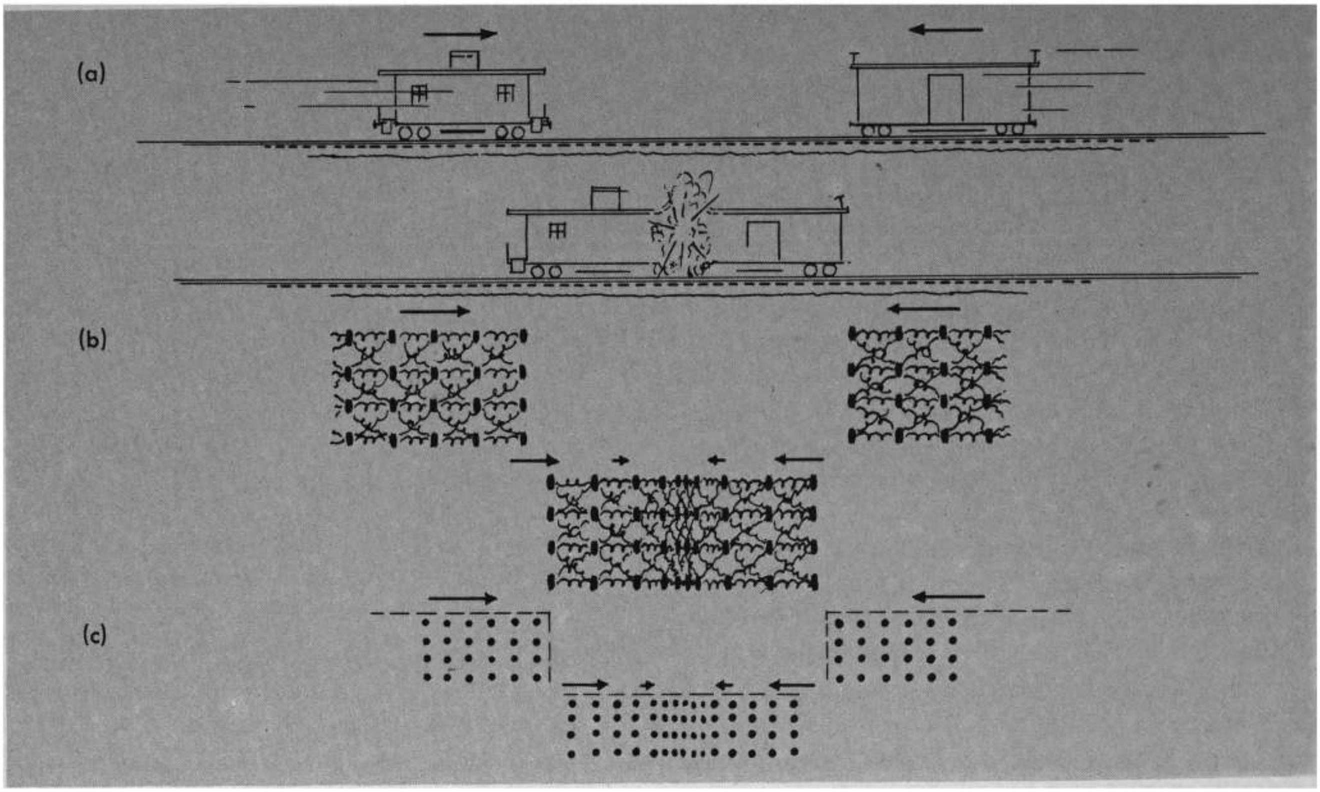
stored when we trapped them in the cage.

Whenever the two masses are held at a given separation, there is a definite potential energy. This does not depend on how fast the system is moving or on how the masses were pushed together. And all this energy will come out as kinetic energy when we let them go. The whole arrangement is very much like the spring bumper—where we could have stored the potential energy by applying a latch to hold the spring compressed. However, this time we have no visible spring. All we have is the forces, and we say that this potential energy is stored in the force field of the interaction. The force field behaves like an imaginary spring.

Actually we should say things the other way around. In a spring the potential energy is really stored in the force fields of interactions between its atoms. The visible shape of the spring merely tells us where its atoms are.

To understand more fully what happens to the atoms, consider what happens when two large masses hit with a bang. As they come into "contact," atoms of one body approach the nearest atoms of the other so closely that large interaction forces arise and energy is stored in the interatomic force fields of both masses. We might picture the collision of two freight cars as compressing interatomic springs [Fig. 18 – 8 (a) and (b)]. A better picture would show the atoms with no real springs, but only their force fields to store energy [Fig. 18 – 8 (c)].

Obviously only some "contact" collisions (for example, those of rubber or steel balls) can store the energy in returnable form. Others leave the atomic systems in a jangle of motion; the large masses are left hotter and move apart with less visible kinetic energy. The collision is inelastic. When we look at one large mass colliding inelastically with another, we see that the forces are not the same on the outward trip after collision as on the inward trip. Some of the visible kinetic energy of the motion of the large masses disappears permanently. (We calculate it by using $\frac{1}{2}mv^2$ for each large mass and do not go into atomic details.) Because this lost kinetic

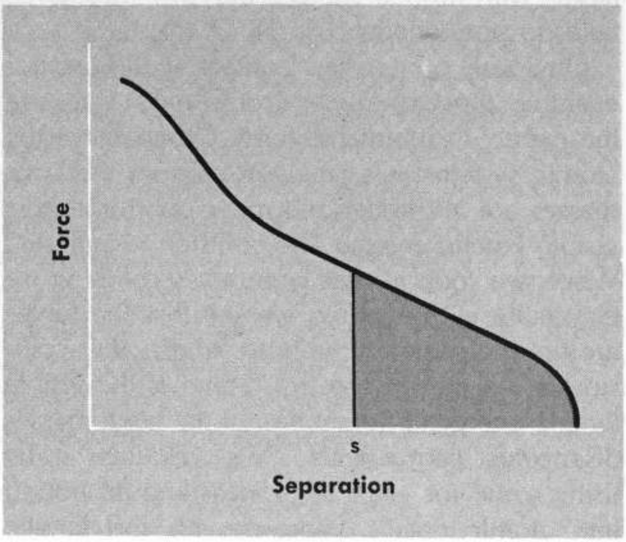


18–8 In (a) two freight cars are colliding. How is the energy stored to make them rebound? We might imagine that the crash compresses a very large number of small springs between the atoms (b). A better picture (c) would show no interatomic springs, but energy is stored in the force fields between the atoms.

energy is not directly returnable, we do not regard it as potential energy.

For elastic collisions, the potential energy depends only on the separation *s* between the masses *A* and *B*. It is given by the work necessary to bring them to this separation from a separation larger than the range of interaction. It is the shaded area under the force-separation curve shown in Fig. 18 – 9. Because the potential energy is only a function of separation, it does not depend on the method of bringing the masses to separation *s*. We have already seen that we can let the masses collide head on with any kinetic energy or push them together by hand. The resulting potential energy is the same.

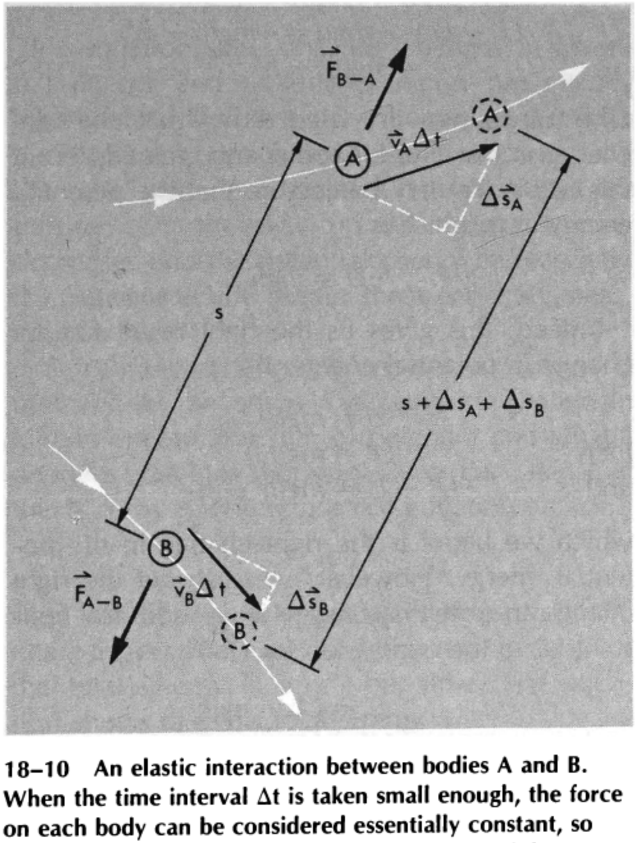
The potential energy is useful because it is a function of the state of the system at the moment and not of its history. Because it is independent of history, we do not need to restrict ourselves to collisions that occur along a fixed straight line. So long as the forces of interaction between the two masses depend only on their separation, there is a definite potential energy



18–9 A graph of force versus separation for two interacting masses. The shaded area represents the work required to bring the masses to separation s. It gives their potential energy when they are s apart.

at a given separation. The masses may be moving in any direction.

Let's look at such a collision in somewhat more detail. The motion of each mass during a short time interval Δt can be split into two components, one along the line joining them and one perpendicularly across it (Fig. 18 – 10). One of these components is in the direction of the force acting on the mass and the other is perpendicular to it. From Chapter 17 we know that when force and motion are perpendicular to each other, no work is done. So the components of the motions perpendicular to the forces are not involved in any transfer of energy. On the other hand, when force and motion are in the same direction work is done, energy is transferred. Consequently (since the forces are equal in magnitude and along the displacements $\Delta \vec{s}_A$ and $\Delta \vec{s}_B$), the magnitude of the forces of interaction times the net change in separation measures the transfer from potential to kinetic energy just as it does for a head-on collision.



18–10 An elastic interaction between bodies A and B. When the time interval Δt is taken small enough, the force on each body can be considered essentially constant, so the work done on it is F times the component of the displacement in the direction of the force. This equals the potential energy given up by the force field and transferred to the body as kinetic energy.

We can apply the same treatment to any isolated system of masses, however complex, so long as all the forces are Newtonian forces—that is, so long as the forces of interaction between each pair of masses are equal and opposite and depend only on the separation. Under these conditions, the potential energy can be defined and the total mechanical energy (kinetic plus potential) of an isolated system remains constant.

We can now see that potential energy may be related to the energy stored in fuels: in gasoline or in the nuclei of atoms, for example. In using fuels we do not release the energy by removing a cage; but under appropriate conditions, we can turn at least a substantial fraction of the energy stored in fuel into the kinetic energy of the motions of large masses—or into the kinetic energy of the random motions of molecules. By using the concepts of kinetic and potential energy, we can understand many of the complex transformations of energy that go on in nature.

A rough analogy may help us to understand kinetic and potential energy. When one man pays another a certain number of dollar bills, the first man is poorer by that number of dollar bills and the second man is richer by the same number. The transfer of dollar bills from one man to the other is analogous to the transfer of kinetic energy from one body to another.

Now suppose that the two men have a shoe box labelled "money" into which they can put dollar bills and out of which they can take them. The first man puts some of the bills into the money box, and the bills stay there for some time before the second man takes them out. During the time the bills are in the box, the total number of bills in the pockets of the two men is less. But the other bills are in the box, and the men can take them out later. The store of bills in the box may be compared to potential energy, which acts as a store of energy that can later be taken out as kinetic energy.

The first man might put bills into the box faster than the second man takes them out. The number of bills in the money box then increases. Similarly, one mass interacting with another may lose kinetic energy faster than the second mass takes it up. The potential energy then increases. In neither case is anything lost;

the bills are in the box and can be taken out; the potential energy is still available. The total remains the same.

18–3 GRAVITATIONAL POTENTIAL ENERGY NEAR THE SURFACE OF THE EARTH

Let us examine the potential energy associated with the gravitational attraction between the earth and other objects. Consider the system formed by the earth and by a body whose mass m is very small compared with the mass M of the earth. If we release the body above the earth's surface, it will fall. Simultaneously the earth will rise (slightly) toward the body. Each mass gains kinetic energy while the potential energy of the system decreases.

At any one instant of time the momenta of the body and of the earth are equal and opposite. Therefore the speed v of the body and the speed V of the earth satisfy the equation

$$mv = MV \quad \text{or} \quad V = \frac{m}{M}v$$

From this relation we can compute the ratio between the kinetic energies of the earth and the body. For the kinetic energy of the earth E_{Ke} we get

$$E_{\mathrm{Ke}} = \tfrac{1}{2}MV^2 = \tfrac{1}{2}M\left(\frac{m}{M}v\right)^2 = \frac{m}{M}\,(\tfrac{1}{2}mv^2)$$

Because $(\frac{1}{2}mv^2)$ is the kinetic energy E_{Km} of the mass m, we see that the ratio between the kinetic energy of the earth and the kinetic energy of the small falling mass is always

$$\frac{E_{\mathrm{Ke}}}{E_{\mathrm{Km}}} = \frac{m}{M}$$

Therefore, if m is very small compared with M, the kinetic energy of the earth can be neglected compared with that of the body. As m falls toward the earth, practically all the potential energy of the system composed of the earth and the mass m is transferred into the kinetic energy of the motion of m.

Now assume that m moves through a distance which is small compared with the radius of the earth. This is certainly true in any laboratory experiment. Then the force of attraction between the earth and the body has a constant magnitude mg. It also has a constant direction (vertically down).

Under the influence of this force a falling body gains speed with a constant downward acceleration g. As we know from our previous work (Section 11 – 8), if the speed increases from v to v' as the body moves downward a distance d, then

$$v'^2 - v^2 = 2gd$$

and the change in kinetic energy is

$$E'_{\mathrm{Km}} - E_{\mathrm{Km}} = \frac{m}{2}(v'^2 - v^2) = mgd$$

The left side of this equation is the change in kinetic energy, while the right side is the work (force, times distance) which measures the transfer of potential energy *into* kinetic energy when the separation between the small mass and the center of the earth *decreases* by the distance d.

In falling from height h above the surface of the earth to height h' a body moves down the distance $d = h - h'$, and the potential energy changes by

$$\begin{aligned} U' - U &= -mgd = -mg(h - h') \\ &= mg(h' - h) \end{aligned}$$

(The minus sign in $-mgd$ shows that potential energy is lost and kinetic energy gained.) From this result we may conjecture that the potential energy at height h is

$$U = mgh$$

Indeed, this gives us the right result for the change in potential energy. If

$$\begin{aligned} & U = mgh \\ \text{then} \quad & U' = mgh' \\ \text{and} \quad & U' - U = mg(h' - h) \end{aligned}$$

which we know is the right change in the potential energy. However, we also get the right change in potential energy if we add any constant U_0 to the expression for U. That is, if

$$\begin{aligned} & U = U_0 + mgh \\ \text{and} \quad & U' = U_0 + mgh' \\ \text{then} \quad & U' - U = mg(h' - h) \end{aligned}$$

The value of U_0 makes no difference to our calculation of the change in potential energy.

In physical problems we deal only with changes in potential energy, and the value of

U_0 can be chosen at our convenience. When we are working near the surface of the earth we often choose $U_0 = 0$. This gives the potential energy the value zero at the earth's surface. Even so, we often have to make a choice of what we shall call the earth's surface. Is it the top of a building, the surface of a street, the bottom of a hole, sea level (Fig. 18 - 11)? We can make any choice which is convenient. Furthermore, when we work with satellites that can go far from the earth we usually choose the potential energy zero infinitely far from the earth. Basically the choice does not matter.

In the equation

$$U' - U = mg(h' - h) = -mgd$$

mgd is the downward force mg on the mass m times the distance d it falls. This expression, therefore, gives the change in the kinetic energy as the mass falls from h to h'; this is

$$mgd = E_K' - E_K$$

Consequently

$$U' - U = -mgd = -(E_K' - E_K)$$

This equation states that the changes in potential energy and in kinetic energy are exactly opposite. As the mass falls, the potential energy decreases and the kinetic energy increases by the same amount. The work mgd measures the transfer. On the way up, the kinetic energy decreases and the potential energy increases by the same amount. Again the transfer is measured by mgd.

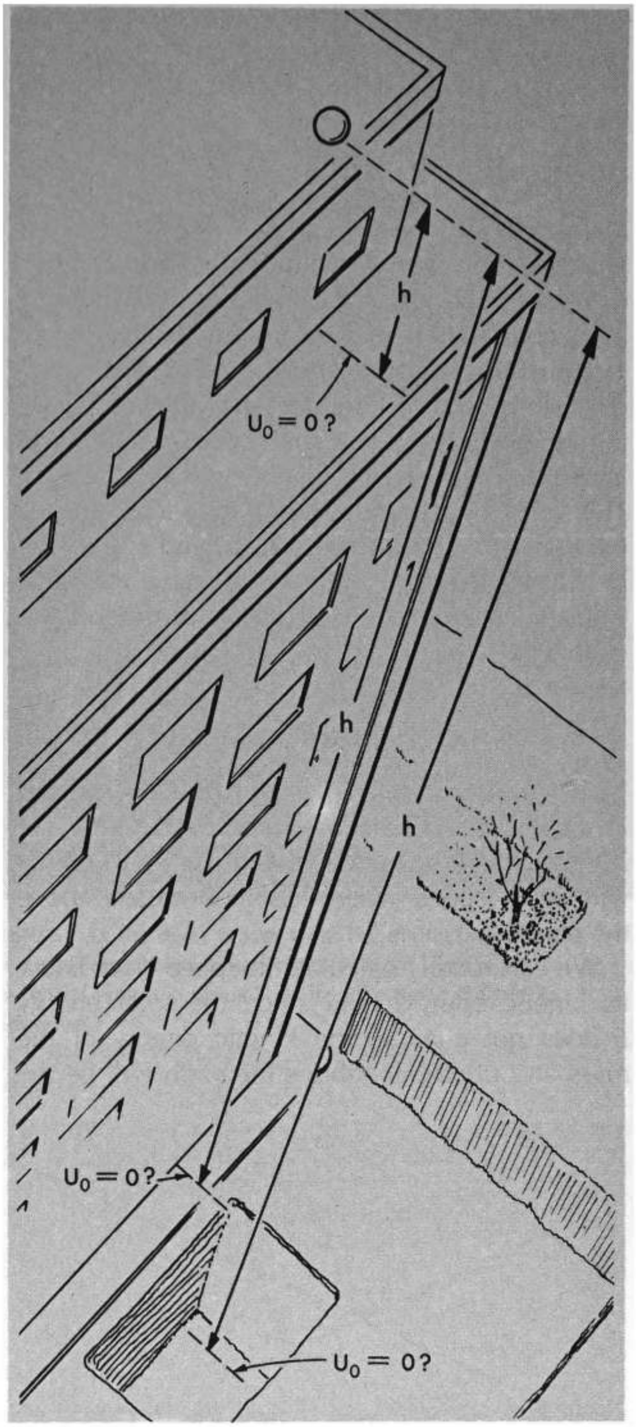
Since any change in potential energy is matched by an equal and opposite change in kinetic energy, the sum of potential and kinetic energies remains constant. We can see that this is so by rewriting our last equation as

$$U + E_K = U' + E_K'$$

The left side of this equation gives the total energy E at one time, and the right side gives the total energy at any other time. The equation shows that this total energy

$$E = U + E_K$$

is the same at any two different times. It is constant, although the values of U and E_K may change. This conservation law applies to the system when it is isolated from external influences which might do work on it and thus change the total energy. It should be clearly understood that no other forces are acting. The relationship between total energy, potential


h
$U_0 = 0$?
h
h
$U_0 = 0$?
$U_0 = 0$?

18-11 We may choose any height we want for the location of zero potential energy. The choice is arbitrary and makes no difference, since we deal only with changes in potential energy.

energy, and kinetic energy is shown in Fig. 18 – 12.

If we choose $U_0 = 0$, we can write the total energy as

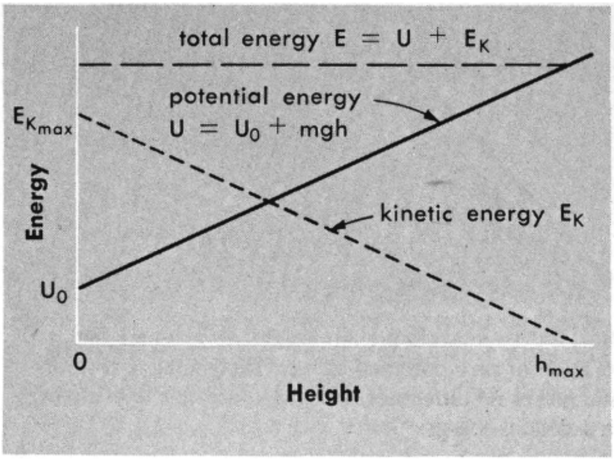
$$\begin{aligned} E &= U + E_K \\ &= mgh + \tfrac{1}{2}mv^2 \end{aligned}$$

This equation is often valuable because it gives us information about the speed of m at different places without reference to the details of the motion from one place to the other. For example, when a projectile moving upward passes the height h, it has the same speed v that it will later have at the same height on its way down. This must be true because mgh is the same on the way up as on the way down, and E is always the same. Therefore, $\frac{1}{2}mv^2$ must have the same value even though the direction of motion has changed.

As a specific application of

$$E = mgh + \tfrac{1}{2}mv^2$$

let us suppose that a mass of 1.00 kg is moving in any direction with a speed of 1.00 m/sec at a height of 3.00 meters above the earth. What will be its speed when it is 2.00 meters above the earth?



18–12 A graph of the potential, kinetic, and total energies for a mass m moving near the earth's surface. Gravity is the only force acting. The potential energy of the earth-mass system is chosen as U_0 at $h = 0$. When the total energy has the value shown, the mass can rise to a maximum height h_{max}, at which point it has no kinetic energy.

We can readily calculate its speed if we know its kinetic energy. And since the total energy E does not change, the kinetic energy of the mass at 2.00 meters above the earth will be the difference between the total energy and the potential energy at 2.00 meters. The total energy is the sum of the initial potential energy (at 3.00 meters height)

$$\begin{aligned} U = mgh &= (1.00\ \text{kg})(9.80\ \text{m/sec}^2)(3.00\ \text{m}) \\ &= 29.4\ \text{joules} \end{aligned}$$

and the initial kinetic energy

$$\begin{aligned} E_K = \tfrac{1}{2}mv^2 &= \tfrac{1}{2}(1.00\ \text{kg})(1.00\ \text{m/sec})^2 \\ &= 0.500\ \text{joule} \end{aligned}$$

That gives

$$E = U + E_K = 29.4 + 0.5 = 29.9\ \text{joules}$$

At 2.00 meters height the potential energy is

$$\begin{aligned} U' = mgh' &= (1.00\ \text{kg})(9.80\ \text{m/sec}^c)(2.00\ \text{m}) \\ &= 19.6\ \text{joules} \end{aligned}$$

and therefore at 2.00 meters height the kinetic energy is

$$\begin{aligned} E_K' = \tfrac{1}{2}mv'^2 = E - U' &= 29.9 - 19.6 \\ &= 10.3\ \text{joules} \end{aligned}$$

When we substitute $m = 1.00$ kg and solve for v' we find

$$v' = 4.54\ \text{m/sec}$$

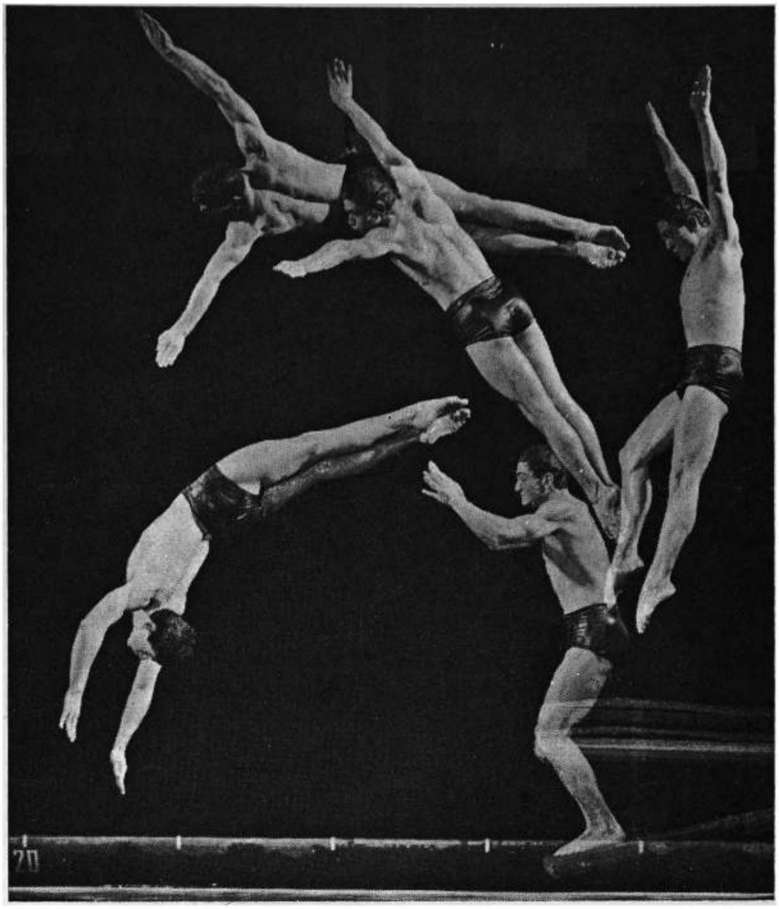
In each of the above motions, the total energy E remains constant so long as the force of gravity is the only force. But we can change the total energy E by pushing the mass, thus doing some work on it. If we then release it, the mass will move with a new constant value of E.

In this section, just as in the example of the mass colliding with a spring bumper, we have been able to find a potential energy which depends only on the position of the mass. This potential energy decreases by exactly the amount that the kinetic energy increases. Just as in the case of the spring bumper, the sum of the potential and kinetic energies remains constant.

Two examples may serve to illustrate the great importance of gravitational potential energy for us. In a pile driver, for example, we increase the gravitational potential energy by raising a large mass. Then by releasing the mass and allowing it to fall freely, we let the gravitational potential energy turn into kinetic energy. This energy drives piles into the earth.

By building dams we can hold water at a greater distance from the center of the earth than that at which it would otherwise stay. By allowing the water to drop from the top of a dam to a lower level, we convert potential energy into other forms—for running mills or producing electrical energy to drive motors or light lamps. Each kilogram of water we allow to fall through 10 meters can do work equal to $1(9.8)10 = 98$ joules. By dropping a kilogram through 10 meters every second, we can keep an ordinary light bulb going.

A standard workhorse* can do 750 joules of work per second. If three of them take shifts, they can light a few light bulbs all year for only a few hundred dollars. The power company will do it for about one tenth that cost by using the potential energy from stored water or coal.



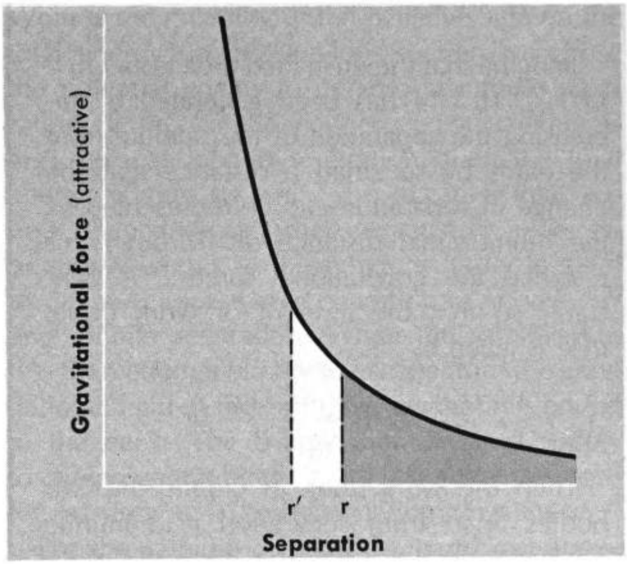
18–13 As the diver moves from the diving board to the water, there are continual changes from kinetic to potential energy and back again. (Courtesy: Harold E. Edgerton.)

18–4 GRAVITATIONAL POTENTIAL ENERGY IN GENERAL

In the last section we determined the gravitational potential energy when a mass m is near the surface of the earth. We assumed that the strength of the gravitational field was constant. On the other hand, we know that when two masses move over larger distances, the gravitational force between them changes. What is the correct expression for the gravitational potential energy in this more general case? What, for example, is the gravitational potential energy of the system consisting of the earth and a satellite?

As we know from Chapter 15, the force of attraction between the earth M and a satellite m is GMm/r^2, where r is the distance to the satellite from the center of the earth. The force separation curve in Fig. 18 – 14 is a graph of this function.

The area under this curve between two different separations measures the work done as the separation changes. For instance, the unshaded area represents the work needed to increase the separation from r' to r. We can find the area under the force-separation curve from any separation r to infinity. This is the dark gray-



18–14 A graph of the force of gravitational attraction versus the distance to the center of the earth. The unshaded area represents the work needed to increase the separation between the earth and a body from r′ to r. The dark-gray area is the work needed to bring the body from separation r to infinity.

---

*The output of workhorses was determined by James Watt in order to make a comparison with the output of his steam engines. More accurately, his standard is 746 joules/sec and is called a horsepower in honor of the horse. A joule/sec is called a watt in honor of Watt.

shaded area extending infinitely to the right from r. It is

$$\frac{GMm}{r}$$

as can be seen from the calculations in the box. This, then, is the work to pull the bodies apart from separation r to infinite separation. This work measures the energy transferred into potential energy of separation to force the bodies apart. The work is therefore equal to the difference $U_\infty - U_r$ between the potential energy at infinite separation and the potential energy at separation r; that is

$$U_\infty - U_r = \frac{GMm}{r}$$

On solving this equation for U_r, we see that the potential energy at the separation r is

$$U_r = U_\infty - \frac{GMm}{r}$$

and if we set the potential energy equal to zero at infinite separation, we get

$$U_r = -\frac{GMm}{r}$$

Derivation of the Expression for Work Done in a Gravitational Field

Imagine that the unshaded area from r to r' in Fig. 18 – 14 has been generated by increasing the separation of the satellite from the earth by so small a distance that the change in separation can be represented by the infinitesimal distance dr. Through this distance the gravitational force F is constant, so that the amount of work done, dW, is given by

$$dW = F dr = \frac{GMm}{r^2} dr$$

Then the work done in pulling the two bodies apart from separation r to infinite separation is

$$W = GMm \int_r^\infty \frac{dr}{r^2} = GMm \left[-\frac{1}{r} \right]_r^\infty = \frac{GMm}{r}$$

Now that we have the potential energy, we can get the total energy E by adding the kinetic energy E_K to it. The result is

$$E = E_K + U_r = E_K - \frac{GMm}{r}$$

Just as we found for masses moved by springs and for objects moving near the earth's surface, the total energy is conserved. For any motion under the influence of gravitational attraction alone, what is gained in potential energy as the bodies separate is paid for by a reduction in kinetic energy. For instance, as a satellite goes around the earth it goes slowest when it is farthest away and fastest when it is nearest the earth. As the earth goes around the sun the same relations are true.

18–5 ESCAPE ENERGY AND BINDING ENERGY

With what energy must we launch a rocket in order for it to escape entirely from the earth? We found in the last section that the gravitational potential energy for a mass m at a separation r from the center of the earth is just

$$U_r = -\frac{GMm}{r}$$

Here M is the mass of the earth, and the zero of potential energy has been chosen at infinite separation. If we wish to launch the rocket from the surface of the earth (where $r = r_e$) with enough energy so that it will just be able to get out to very large separation with no extra energy left over, we must give it in the launching an initial kinetic energy of

$$E_K = +\frac{GMm}{r_e}$$

Then the change in kinetic energy, from $+\frac{GMm}{r_e}$ to zero, just compensates for the change in potential energy, from $-\frac{GMm}{r_e}$ to zero, or $\Delta E_K = -\Delta U$ as we expect. This amount of kinetic energy is the minimum which the rocket can have to just barely escape from the earth's gravitational pull. You will notice that this escape kinetic energy, $E_K = \frac{GMm}{r_e}$, is directly proportional to the mass of the satellite; to blast off a more massive rocket requires more energy, just as you would expect. If we put in numerical values for the gravitational constant G, the mass of the earth M, and the radius of the earth r_e from Chapter 15, we find that the required escape kinetic energy is 6.24×10^7 joules for each kilogram of mass of the rocket. For a rocket having a mass of one metric ton (1000 kg) to escape completely from the earth requires that we give it an energy of 6.24×10^{10} joules.

We can also find the initial speed v_e which the rocket must have in order to escape from the earth. Since kinetic energy is $\frac{1}{2}mv^2$, we can write the escape kinetic energy as

$$E_K = \frac{GMm}{r_e} = \tfrac{1}{2}mv_e^2$$

Since the mass of the rocket appears in both expressions for E_K, we can eliminate it and find

$$\tfrac{1}{2}v_e^2 = \frac{GM}{r_e} = 6.24 \times 10^7 \frac{\text{joules}}{\text{kg}}$$

which is now independent of the rocket mass. [If you work out the units, you will find that 1 joule/kg = 1(m/sec)^2.] Thus

$$v_e = \sqrt{12.48 \times 10^7} = 1.12 \times 10^4 \text{ m/sec}$$

or 11.2 km/sec. The quantity v_e, usually called the "escape velocity," does not depend on the mass which is being launched. It is the initial speed that must be given to *any* body in order for it to escape from the earth and never return.

Many of our most interesting rockets are designed not to escape the earth's gravitational pull, but to place a satellite into a fairly stationary orbit around the earth just a few hundred kilometers above its surface. It is interesting to compare the energy needed to place a satellite in orbit with the energy that would be required to cause it to escape. To simplify the calculation, let us make two fairly good assumptions. First, we shall assume that the orbit is circular—this requires very delicate control of the launching conditions, but it can be done; second, we shall approximate the radius of the orbit by the radius of the earth; if the satellite is just above the earth's atmosphere, as in manned satellites, for example, these radii differ by only two or three percent.

In order to move in a circular orbit of radius r_e at constant speed v, a satellite of mass m requires a centripetal force $\frac{mv^2}{r_e}$ which will be supplied by the gravitational force of the earth $\frac{GMm}{r_e^2}$. Therefore

$$\frac{mv^2}{r_e} = \frac{GMm}{r_e^2} \text{ or } mv^2 = \frac{GMm}{r_e}$$

Thus the kinetic energy of the orbiting satellite is

$$E_K = \tfrac{1}{2}mv^2 = \tfrac{1}{2}\frac{GMm}{r_e}$$

This is just half of the kinetic energy of escape. In other words, to put a satellite into orbit just above the earth's atmosphere requires only half the energy that it takes to throw the satellite permanently away from the earth.

Now, an orbiting satellite is still in the gravitational field of the earth; so, in addition to kinetic energy, there is also a potential energy of separation $U = -\frac{GMm}{r_e}$ at the radius r_e. We can therefore write the *total energy* of the orbiting satellite-earth system as

$$E = E_K + U = \tfrac{1}{2}\frac{GMm}{r_e} + \left(-\frac{GMm}{r_e}\right)$$
$$= -\tfrac{1}{2}\frac{GMm}{r_e}$$

What is the meaning of the negative sign on the total energy? In the previous section we chose the potential energy to be zero at infinite separation. If the kinetic energy of the satellite is also zero at infinite separation, then the total energy is zero as well. The negative total energy, then, indicates that we must do work on or supply energy to the system in order to get it to the zero total energy condition of no velocity and infinite separation. When the satellite has less than enough energy to escape from the gravitational field of the earth, we say that it is bound to the earth. The energy that we must supply to overcome the binding and just allow the satellite to escape is called the "binding energy."

For the satellite orbiting the earth just above the earth's atmosphere, we must supply the energy to bring E up to zero, so

$$\text{binding energy} = +\tfrac{1}{2}\frac{GMm}{r_e}$$

For our metric-ton satellite in such an orbit, the binding energy comes out to be 3.12×10^{10} joules. For a metric-ton satellite which is stationary on the earth's surface before launching, the total energy is just the potential energy $-\frac{GMm}{r_e}$, so the binding energy needed to overcome this is $+\frac{GMm}{r_e} = 6.24 \times 10^{10}$ joules (just twice as much as for the orbiting case). Your own binding energy on the surface of the earth (the energy necessary to send you to outer space) amounts to about 4 or 5×10^9 joules. We can even find the binding energy of the earth to the sun: it is just

$$\text{binding energy} = \frac{GM_{\text{sun}}m_{\text{earth}}}{2r_0}$$

where r_0 is the radius of the orbit of the earth around the sun. The proof is identical to the one we did for the satellite orbiting the earth. Putting in values from Chapter 15, we find that the binding energy of the earth to the sun is a bit more than 2×10^{33} joules! That's a lot of energy!

For the general case, when an object in a gravitational field has a total energy

$$E = \tfrac{1}{2}mv^2 - \frac{GMm}{r}$$

which is negative, the object is "bound" by the gravitational field with a binding energy of

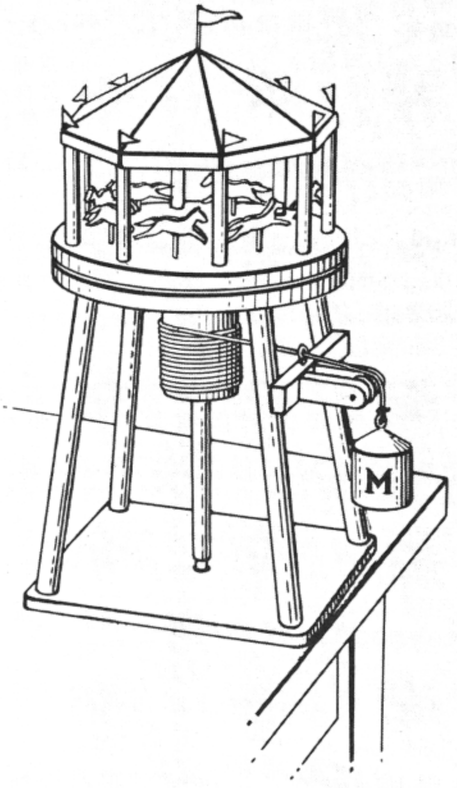
$$-E = \frac{GMm}{r} - \tfrac{1}{2}mv^2$$

There are binding energies of importance in atomic physics also, that of the electron to the proton in hydrogen, for example. However, the scale of these binding energies is quite different, and they arise from electrical rather than from gravitational forces. We shall postpone our discussion of these atomic binding energies until later in the course.

18–6 TOTAL MECHANICAL ENERGY

In this chapter we have investigated the potential energy of systems in which the parts interacted through forces that were equal and opposite and which pointed along a line connecting their centers of mass. The parts acted like point masses; that is, each acted as if its mass were concentrated at its center of mass. Even the gravitational attraction of the earth on a mass nearby appears to come from the point at the earth's center. But it is not clear that all mechanical systems involve only such simple forces of interaction, nor that their parts can be treated as points. What happens to energy in mechanical systems where we cannot be sure that all interactions are like these?

To answer this question, suppose we have a toy merry-go-round (see Fig. 18–15) which is driven by a weight-and-pulley system. Several turns of string are wrapped around the shaft, and then the string goes through a pulley and down to the mass M. If we let the merry-go-round go, it will start to rotate, gradually speeding up as the mass M falls. No one would try to analyze this system by finding a set of interaction forces between all the atoms. But by viewing the merry-go-round as a collection of small masses and finding their speeds, we can calculate the kinetic energy of each piece.



18–15 A toy merry-go-round driven by a falling weight. As the weight falls, the merry-go-round begins to rotate, turning faster and faster. The whirling platform and the falling weight gain kinetic energy which is equal to the loss in potential energy of the weight.

Then, adding up, we can find the total kinetic energy of the whirling merry-go-round. We can also measure the distance that the mass M has fallen and the kinetic energy of the mass. When we do such an experiment with a merry-go-round on nearly frictionless bearings, we find that the kinetic energy gained by the merry-go-round and the mass M is just equal to the decrease in the gravitational potential energy as the mass M falls toward the earth. Eventually the string is unwound, but the whirling merry-go-round continues to turn. The string then wraps around the center post in the other direction. The weight rises, and the merry-go-round slows down. When the merry-go-round stops turning, the mass M is almost back up to its original position. It fails to return to the original height only because of the energy transferred into other forms through frictional interactions.

We can consider more and more complex mechanical systems and, without analyzing the forces of interaction in detail, we can often evaluate the kinetic energy and the potential energy. Their sum then represents the total mechanical energy in the system. When frictional interactions are very small, the energy may change back and forth between potential and kinetic, but the total mechanical energy remains constant.

But sometimes we get into trouble; frictional interactions frequently cannot be made so small that we can neglect them. We then see the total mechanical energy disappear and the parts of the mechanical system get warmer. In the next chapter we shall learn how to take account of this energy which is transferred away as heat. Then we shall see why we believe that energy is always conserved.

QUESTIONS

1.* In Fig. 18 – 1, where is the acceleration of the block the greatest?

2.* How could you determine the k for the spring in Fig. 18 – 5?

3. A mass of 4.0 kg sliding with a velocity of 3.0 m/sec on a frictionless horizontal table collides with a queer kind of spring bumper. The bumper exerts a constant force of 120 newtons on the mass as it moves in (compressing the spring) and the same force on the way out until the spring is back where it was.

   (a) Is this an elastic collision? How do you know?

   (b) What is the kinetic energy at the beginning of the interaction?

   (c) How much is the spring compressed?

   (d) What is the ratio of kinetic energy to potential energy when the spring has been compressed 10 cm?

4. A 3.0-kg mass moving with a velocity of 2.0 m/sec collides with a spring bumper which exerts a force $F = 100x$, where F is the force in newtons and x is the compression in meters.

   (a) Draw a graph of F versus x from $x = 0$ to $x = 0.40$ m.

   (b) What is the potential energy stored in the spring when $x = 0.10$ m? What is the kinetic energy of the mass at this point?

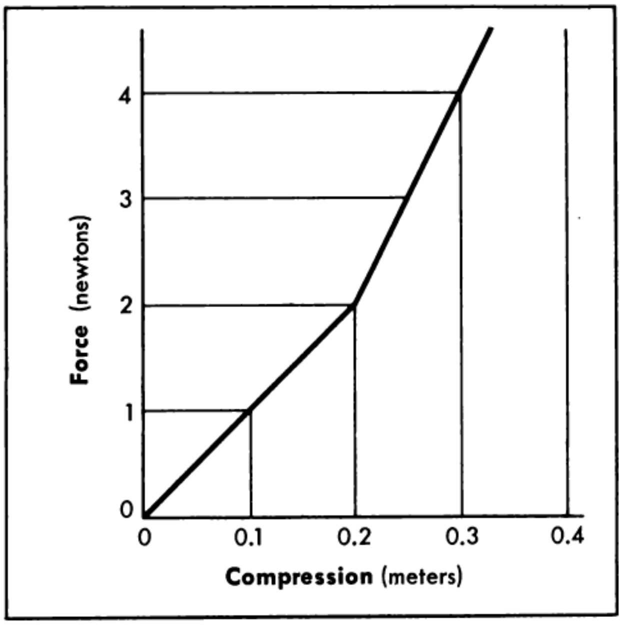
   (c) What will happen if the spring is compressed 0.10 m by hand and then the 3-kg mass is placed in contact with the spring and the hand is removed?

5. The force-compression curve of a spring is shown in Fig. 18 – 16.

   (a) How much work is done in compressing the spring 0.3 m?

   (b) What is the potential energy of the spring when compressed this amount?

   (c) Place a 2-kg mass at rest against the spring when it is compressed 0.3 m. Let go. What is the kinetic energy of the mass as it passes the point where the spring is compressed 0.2 m?



18–16 For Question 5.

6. Pretend you are given two springs of different sizes. How would you set them up so that the force-compression curve of the system looks like Fig. 18 – 16?

7. A 1.0-kg mass moving with a velocity of 10 m/sec strikes a spring, compressing it a distance of 0.20 m. The mass rebounds with a speed of 8.0 m/sec.

   (a) What is the loss in kinetic energy of the mass?

   (b) What happens to this lost energy?

   (c) Which of the force-compression curves in Fig. 18 – 17 do you think is most likely correct for this spring?

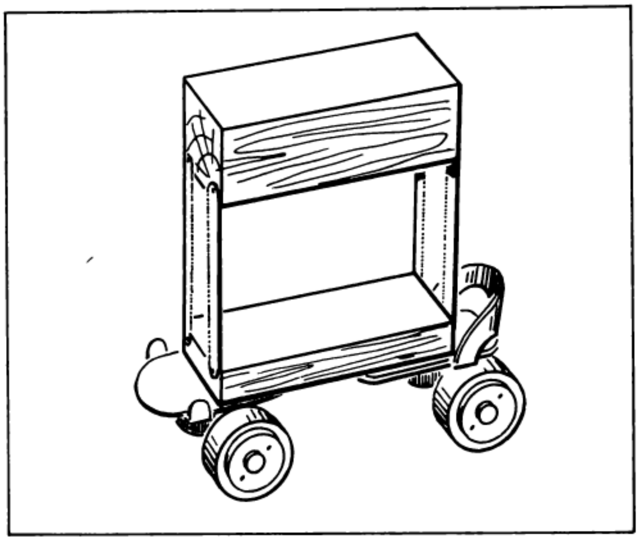
8. A linear elastic spring is compressed 0.2 m by a force of 20 newtons.

   (a) What is the force constant *k* (or force-compression ratio) of the spring?

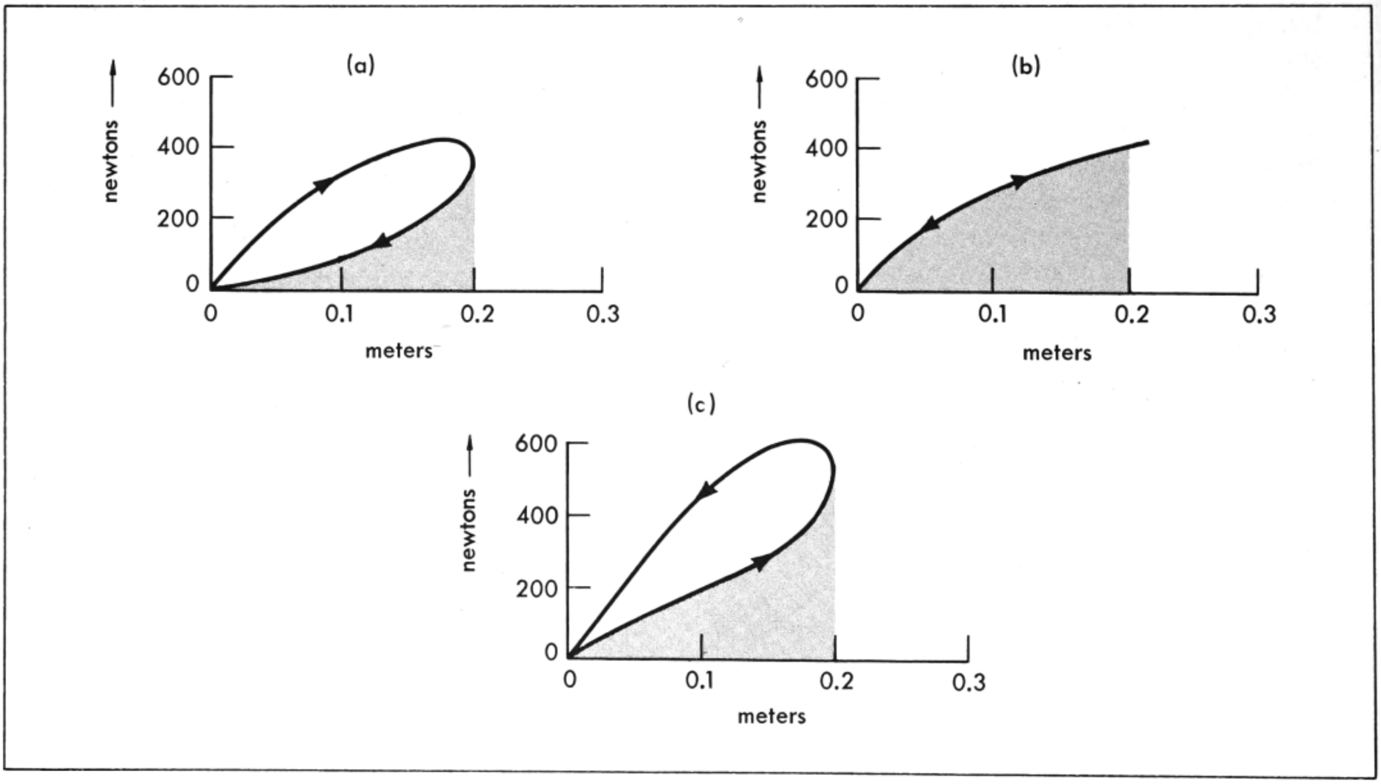
   (b) What is the equation for the potential energy stored by the spring as a function of its compression?

9.* In Fig. 18 – 10, we see two masses which have a certain amount of stored energy at separation *s* and are moving to the right. If we placed the two masses at rest in the same positions and released them, which way would they go?

10. Figure 18 – 18 shows a roller skate with a mass *M* mounted on it by four hacksaw blades attached to a base. The mass *M* is equal to the mass of the roller skate plus the mass of the base. If you hold the roller skate and pull the mass *M* out to the left, and then let the whole system go, the mass moves to the right and the skate to the left. Be prepared to describe the further motion, indicating the form of the energy at various stages.



18–18 For Question 10.



18–17 For Question 7.

11. What will happen if the 3.0-kg mass in Question 4 hooks itself onto the spring and remains fastened to it?

12. In Fig. 18 – 7, why does the large ball not gain kinetic energy while the force of the cage acts on it?

13. Suppose we have a spring-operated popgun something like the one in Fig. 18 – 19. It has a piston of mass m fastened to the end of a light spring whose force constant is k, and it shoots a ball of mass M. When the gun is loaded, the piston is pushed down the barrel until the spring is compressed through a distance d, and latched to the trigger.

    Leaving out friction for the sake of simplicity, work out an argument to decide whether all the potential energy of the spring will be transferred to the ball.

14. A linear spring with constant k is compressed between two Dry Ice pucks of mass m_1 and m_2. The spring is compressed a distance x and then tied with a thread. Both pucks are initially at rest.

    (a) If the thread is burned, what is the total kinetic energy of both pucks?

    (b) What are the momenta of the pucks, and how are they related?

    (c) What is the kinetic energy of each puck in terms of its momentum and its mass?

    (d) What fraction of the total energy does each puck acquire? What is their ratio?

15. A spring bumper with restoring force $F = 200x$, where F is in newtons and x in meters, is compressed 0.100 meter. A 0.500-kg mass is placed next to the end of the spring and the whole thing let go.

    (a) With what momentum will the mass leave the spring?

    (b) We do the same thing with masses 0.125 kg, 2.00 kg, and 8.00 kg. What is the momentum of each as it leaves the spring?

    (c) What is the energy of each as it leaves the spring?

16. Figure 18 – 20 is the graph of the displacements as a function of time of two interacting masses $m_1 = 3$ kg and $m_2 = 1$ kg.

    (a) What is the initial kinetic energy of m_1?

    (b) What is the final kinetic energy of m_1?

    (c) What is the final kinetic energy of m_2?

    (d) What is the minimum total kinetic energy?

    (e) What is the maximum potential energy?

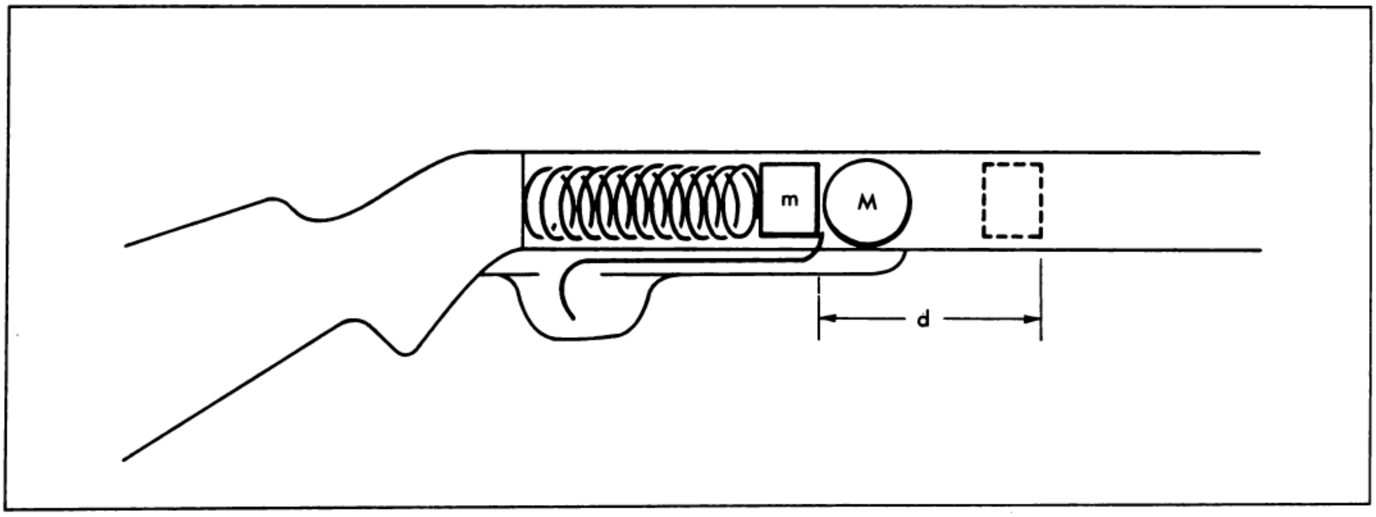
17.* A 70-kg boy climbs 5 meters up a rope. What is the increase of his potential energy?

18. (a) A ball of mass 0.25 kg is thrown to the right at a speed of 2.0 meters per second. It starts along the frictionless surface at the left of Fig. 18 – 21. How high up the slope at the right will it go before coming momentarily to rest? What kind of motion will it perform?

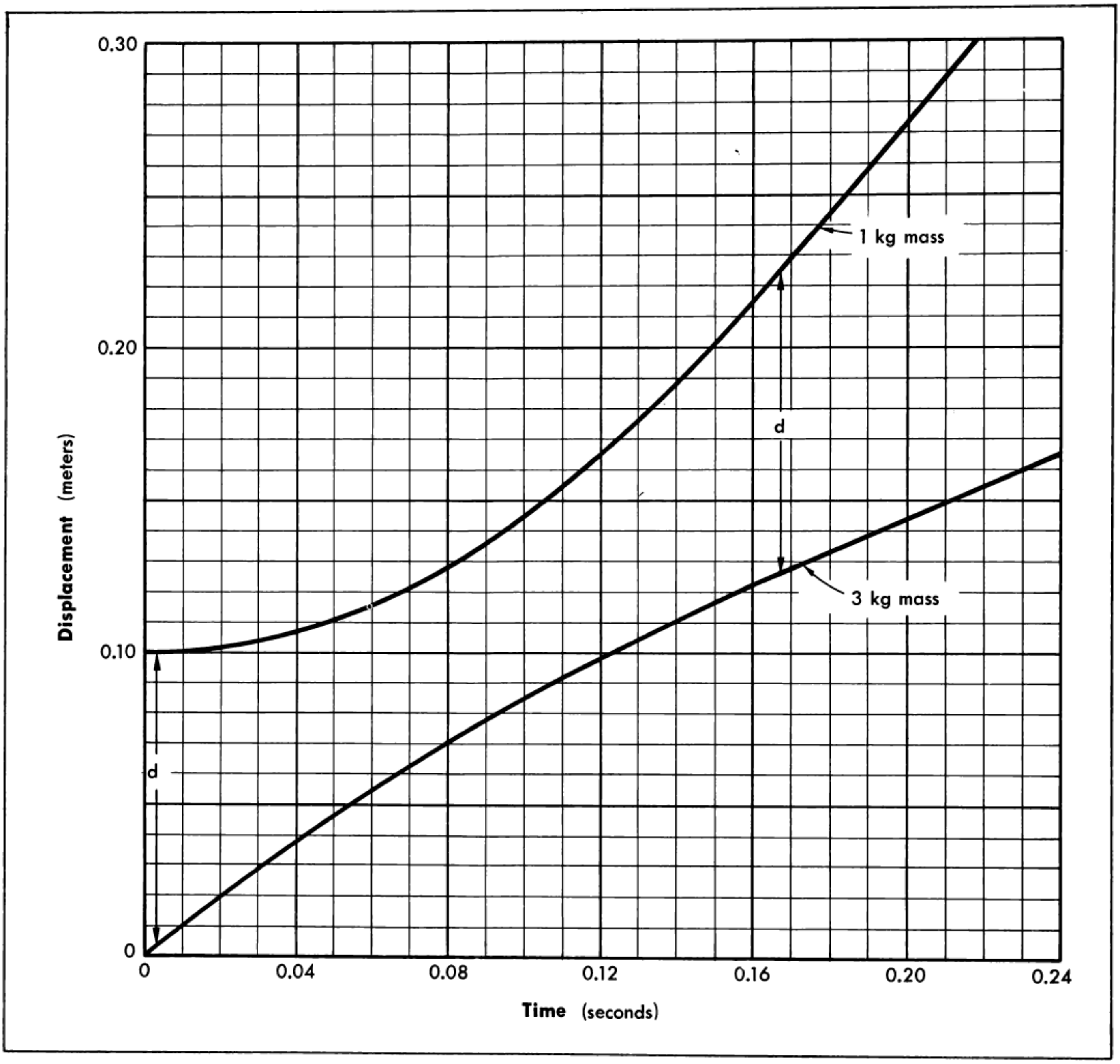
    (b) If the ball is released from rest at point P, what kind of motion will it perform? How high up the slope on the right will it rise? What is the binding energy (the extra energy needed to make the ball escape from the well in the center of the figure)?

    (c) What is the binding energy in part (a)?

19. A linear spring whose force-extension ratio k is 40 newton/m hangs vertically, supporting a



18–19 For Question 13.



18–20 For Question 16.

0.80-kg mass at rest. The mass is then pulled down a distance of 0.15 m.

(a) How high will it rise?

(b) What will be its maximum velocity?

(c) How would the answers to (a) and (b) differ if the experiment were done on the moon?

20. A 0.200-kg stone is thrown upward from a point 20.0 meters above the earth's surface at an angle of 60 degrees with the horizontal and with a speed of 20.0 m/sec.

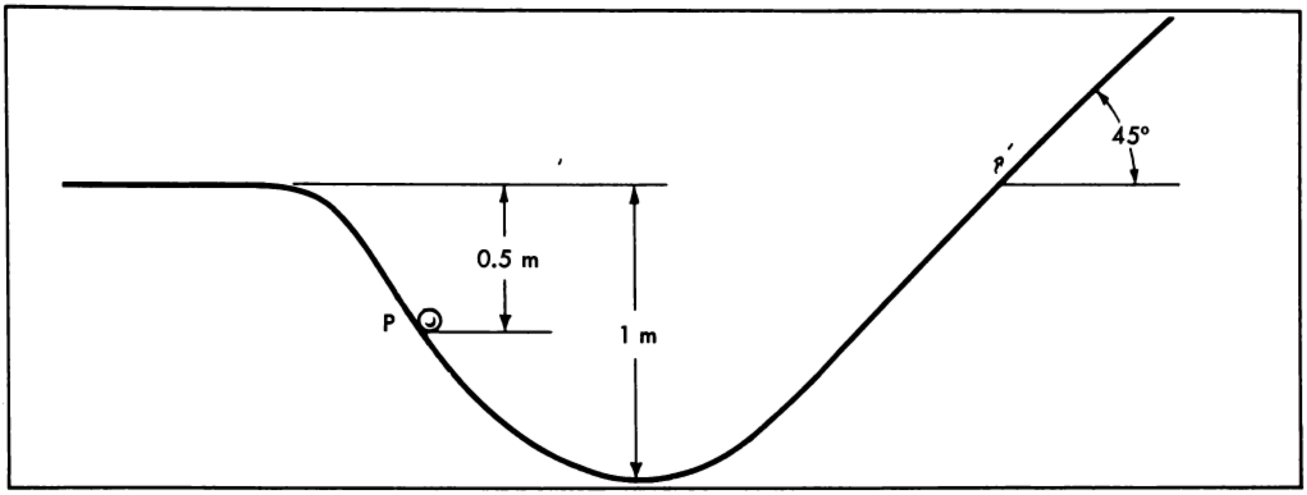
(a) What is its total energy?

(b) What will be its total energy when it is 15.0 m above the earth's surface?

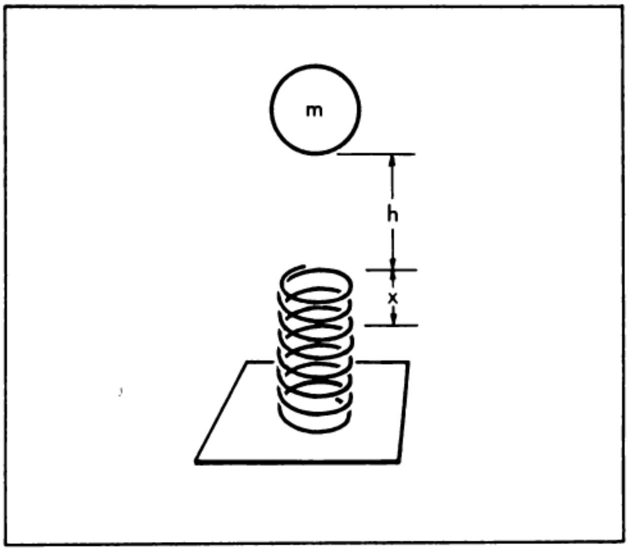
(c) What will be its speed 15.0 m above the earth?

21. A ball of mass m drops from a height h, as shown in Fig. 18 – 22, and compresses the spring, of force constant k, a distance x. The mass of the spring is negligible compared with the mass of the ball.

(a) Express the maximum compression of the spring, x, in terms of m, h, and k.



18–21 **For Question 18.**



18–22 **For Question 21.**

(b) Evaluate x if $m = 4$ kg, $h = 3$ meters, and $k = 500$ newtons/meter.

22. When an object of mass m moves from r to r' in the earth's gravitational field, the potential energy changes by $\Delta U = -GMm\left(\frac{1}{r'} - \frac{1}{r}\right)$ where M is the mass of the earth.

Show that, if the object moves away from the earth a distance $\Delta r = (r' - r)$ which is very small compared with its distance from the earth's center, the above expression reduces to

$$\Delta U = mg\Delta r$$

23. The force field between a pair of protons is repulsive. Does the potential energy increase or decrease as a pair of protons are brought together?

24.* A satellite in a circular orbit about the earth has a kinetic energy E_K. How much more energy would be required to make it escape from the earth?

25.* A rocket of mass m ceases to climb when it has reached a distance $10r_e$ from the center of the earth. How much more energy would it require in order to escape?

26.* How do the escape speeds of a 10-mg grain of sand and a 10^3-kg rocket compare? Ignore the effects of air resistance.

27. What is the binding energy of the proton and electron in a hydrogen atom if the proton and electron are 0.50×10^{-10} m apart and the force of attraction between them is given by

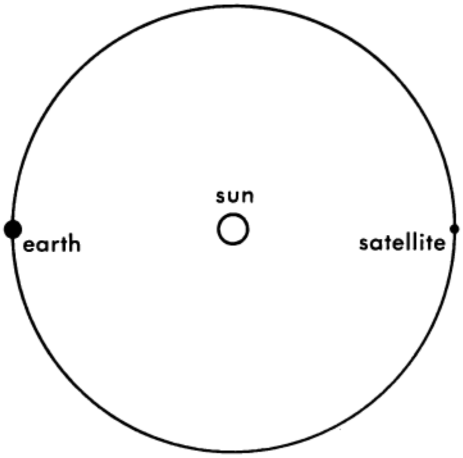
$$F = \frac{2.3 \times 10^{-28}}{r^2} \text{ newtons}$$

when r is in meters?

Remember, the electron will not stand still long with a force on it. Assume it is moving in a circle around the proton.

28. Find the binding energy, to two significant figures, of
    (a) a 70-kg man to the earth.
    (b) the moon to the earth.

29. Would the kinetic energy you would have at the equator due to the rotation of the earth make much difference in your binding energy to the earth?



18–23 For Question 30.

30. (a) How much work would be required to launch a 1-kg satellite from the earth and cause it to move in the earth's orbit but on the opposite side of the sun, as shown in Fig. 18 – 23?

(b) What work would be required for this satellite to leave the solar system from this point?

(c) What is the ratio of the energy required to escape the solar system compared with the energy required to escape the earth, for a satellite launched from the earth?

31.* You are given a very complicated mechanism with levers and wheels in it. It is known that there is no motor inside and that there is very little friction. A string comes out of each side of the apparatus. In order to lift a body 0.10 m on the right-hand string, you must pull 6.0 m on the left-hand string. With what force do you have to pull the left-hand string in order to lift a 12-newton weight on the right-hand string?

32. A drainpipe pointing slightly downhill sticks out from a retaining wall at the side of a road. A ball thrown up the pipe comes back with a speed greater than that with which it was thrown. If you observed this, would you be surprised? What would you suspect?

33. A laboratory cart fitted with a spring bumper is rolled across the table with a velocity v_0 and collides with another cart of equal mass which is standing at rest. The spring constant of the bumper is k. Assuming that friction in the wheels and bumper is low enough that mechanical energy will be conserved, express the compression of the spring at the moment when the two carts are closest together in terms of v_0, m, and k. Work this problem in center-of-mass coordinates. That is to say, consider all the motions in a frame of reference which moves so that in it the total momentum of the two carts is always zero.

19

Angular Momentum and Its Conservation

19–1 THE LAW OF EQUAL AREAS

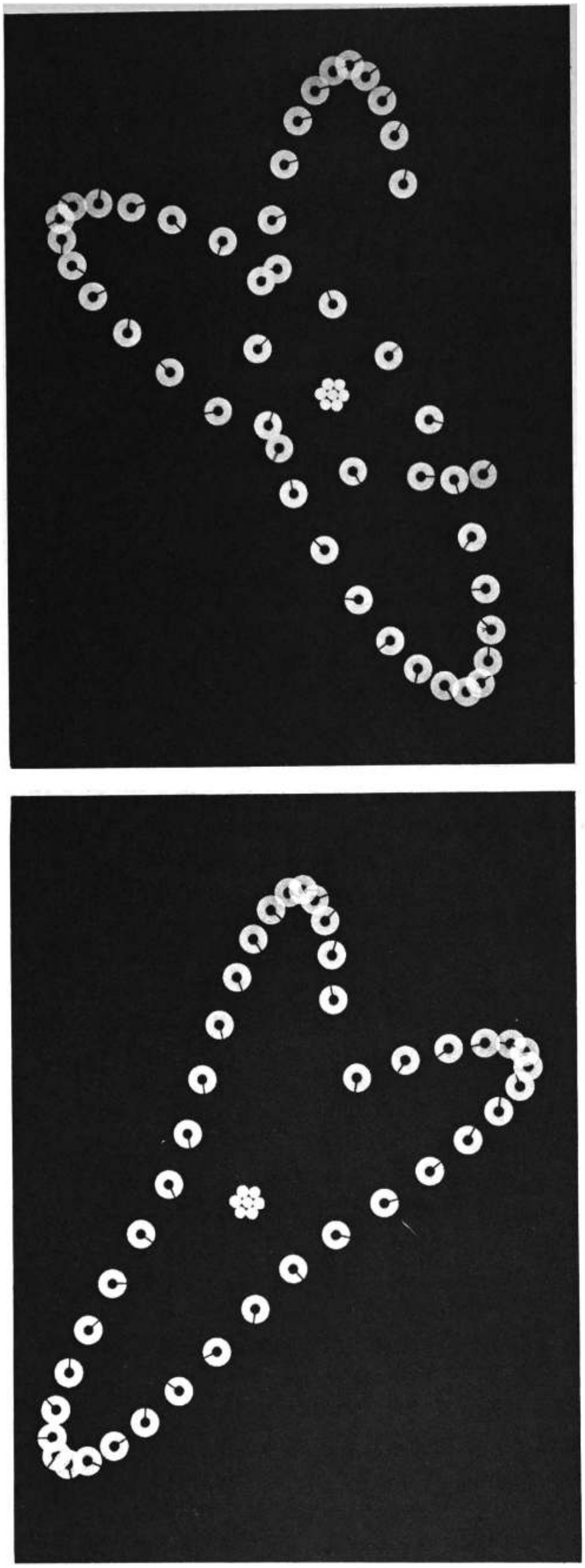
WE HAVE studied several examples of rotational motion: uniform circular motion, the elliptical motion of the planets around the sun, and the motion of a pendulum swinging in an elliptical path. The forces acting on the moving objects, though different, have one property in common; they are all *central forces*—that is, each is directed to a fixed point in space. In our examples these points are respectively the center of the circle, one of the foci of the ellipse, and the center of the ellipse. In all these cases we found that Kepler's law of equal areas holds. That is, in equal time intervals the areas swept out by the position vector pointing from the fixed force center to the moving object are the same.

Figure 19 – 1 shows two examples of the motion of a bar magnet hung from the ceiling in the field of a fixed bar magnet. The details of the arrangement are shown in Fig. 19 – 2. By making the necessary measurements on the photographs, you can convince yourself that the law of equal areas holds also for these cases.

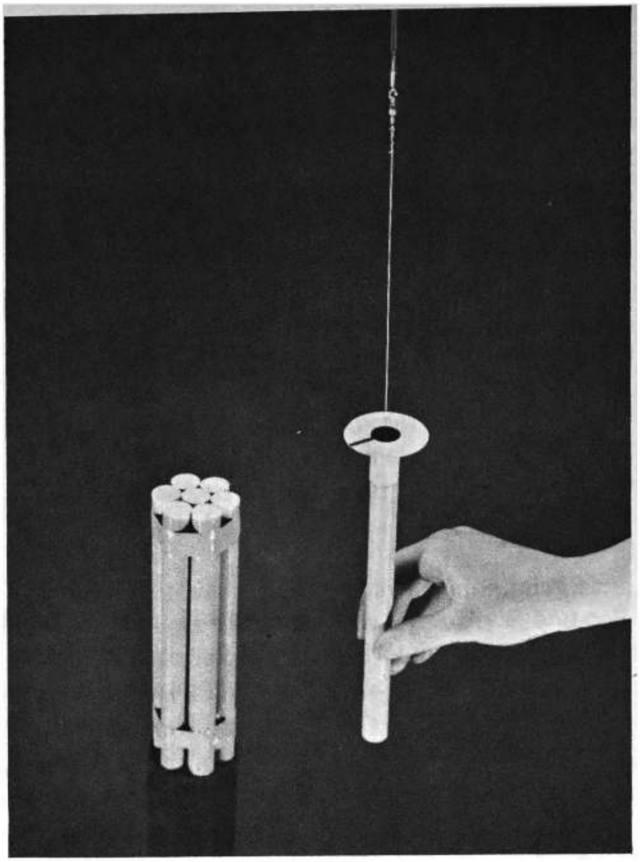
It is quite natural to suspect that the law of equal areas holds for all motions under any central force. Indeed, we can show that it follows directly from Newton's law of motion. To do this, let us consider first the motion of a body on which a force acts intermittently for very short times. We can imagine a moving body which gets a sharp push, say, every second. All the pushes are directed to a fixed center, but they may have any magnitude.

A possible trajectory traversed by the body under these conditions is shown in Fig. 19 – 3. The body starts at *A* and moves with constant velocity to *B* in one unit of time. At *B* it is pushed toward *0* and therefore moves with a different (constant) velocity to *C*. At *C* it gets another push toward *0* and therefore moves on to *D*. Since the blows are delivered at equal time intervals, we can choose our velocity scale so that the vectors $\overrightarrow{AB}$, $\overrightarrow{BC}$, etc., represent the velocity of the object in the appropriate units. The vectors shown in dotted lines are the changes in velocity $\Delta\vec{v}$ caused by the impulse of the forces acting at *B*, *C*, etc.

Because the impulse delivered to the body at *B* is in the direction



19–1 Stroboscopic photographs of the motion of a bar magnet hung from a high ceiling in the field of a fixed magnet. See details in Fig. 19–2.

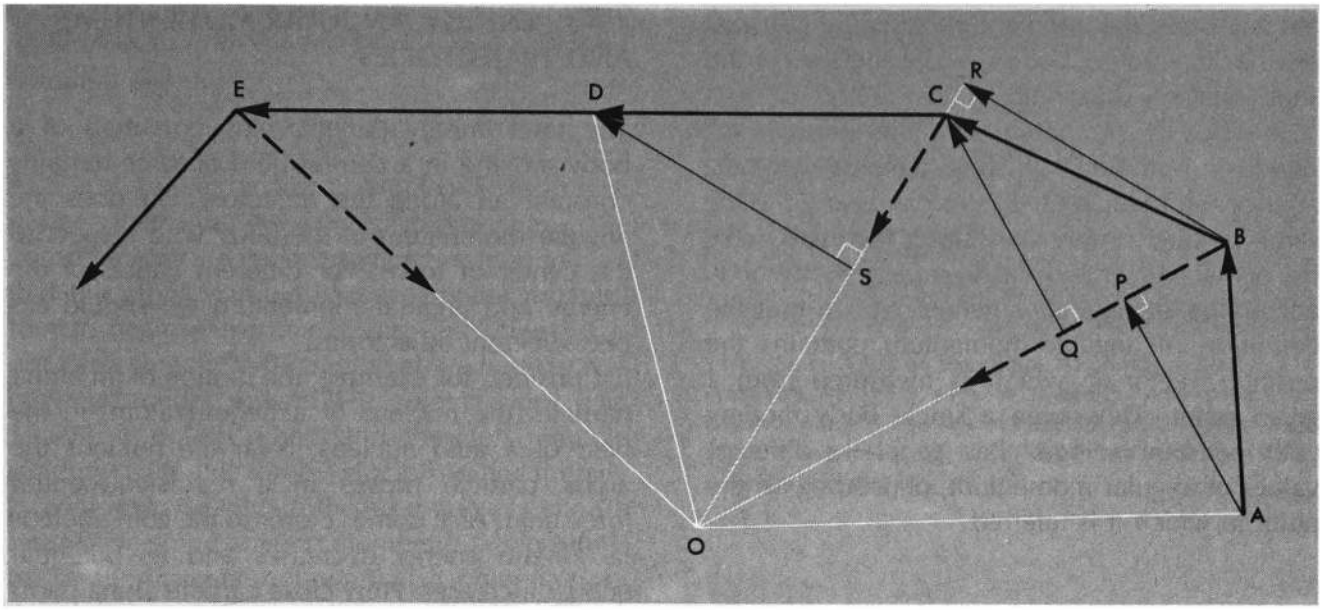


19–2 A close-up of the magnets used in Fig. 19–1.

$\overrightarrow{BO}$, it leaves the vector component of the velocity $\overrightarrow{AB}$ perpendicular to $\overrightarrow{BO}$ unchanged. That is, the vector component $\overrightarrow{AP}$ of $\overrightarrow{AB}$ before the impulse is the same as the vector component $\overrightarrow{QC}$ or $\overrightarrow{BC}$ after the impulse. Similarly, the vector component $\overrightarrow{BR}$ of the velocity $\overrightarrow{BC}$ before the impulse at C equals the vector component $\overrightarrow{SD}$ after that impulse.

The areas swept out by the displacement vector per unit time are equal to the areas of the triangles OAB, OBC, OCD, etc. Let us look first at the triangles OAB and OBC. They have the base OB in common, and we have just seen that their heights AP and QC are equal; hence the two triangles have equal areas. Looking now at the triangles OBC and OCD, we see that they have the base OC in common and the heights $BR = SD$, and hence $OCD = OBC = OAB$. The same reasoning shows that all the successive triangles have equal areas. We see, therefore, that the law of equal areas holds.

The crucial step in our proof is based on the central nature of the successive impulses, all of which are directed toward the point O. How



19–3 Motion of a body subject to several central impulses. The changes in velocity resulting from the impulses are shown by the heavy broken lines. At D the impulse was zero.

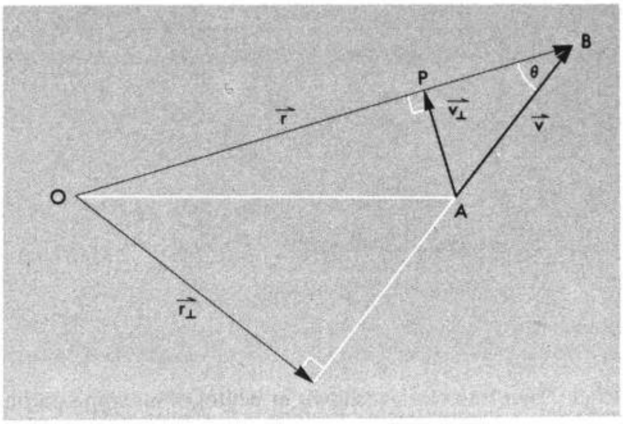
many impulses are delivered per second, or how large the impulses are, doesn't matter. The impulses at *B*, *C*, and *D*, for example, are all different; the one at *D* happens to be zero. It is the central property of the intermittent force which assures us that the vector component of the velocity perpendicular to the line from the center is unchanged.

Now, in reality we have forces which act continuously and make the body move along a curved path. But we can always approximate such a path by one composed of many small straight segments and apply our proof to this angular path. Making the straight segments smaller and smaller will bring us as close to the real path as we wish.

19–2 ANGULAR MOMENTUM

We can look at the law of equal areas as an extension of Galileo's law of inertia: in the absence of any (unbalanced) force the velocity of an object remains constant. In the presence of a central force the velocity changes, but the product of the displacement vector from the center of force times the perpendicular component of the velocity remains constant: $rv_\perp =$ constant. (This constant is twice the area covered per unit time.) The mass of the moving body does not appear in either of these laws.

We use quantities like v and $rv_\perp$ when we describe the motion of a body. When we wish to relate the motion to forces, the mass becomes important. We introduced the momentum $\vec{\mathbf{p}} = m\vec{\mathbf{v}}$ and found that for interacting bodies the total momentum is conserved while the sum of the velocities is not. Similarly we introduce the quantity $L = rp_\perp = mrv_\perp$, which we call the *angular momentum.** We shall see shortly (Section 19 – 6) that under suitable circumstances



19–4 $rv_\perp = r_\perp v = rv \sin\theta$ = twice the area of OAB.

---

*Sometimes the momentum $\vec{\mathbf{p}}$ is referred to as *linear momentum,* to distinguish it from angular momentum. We shall continue to call it simply momentum.

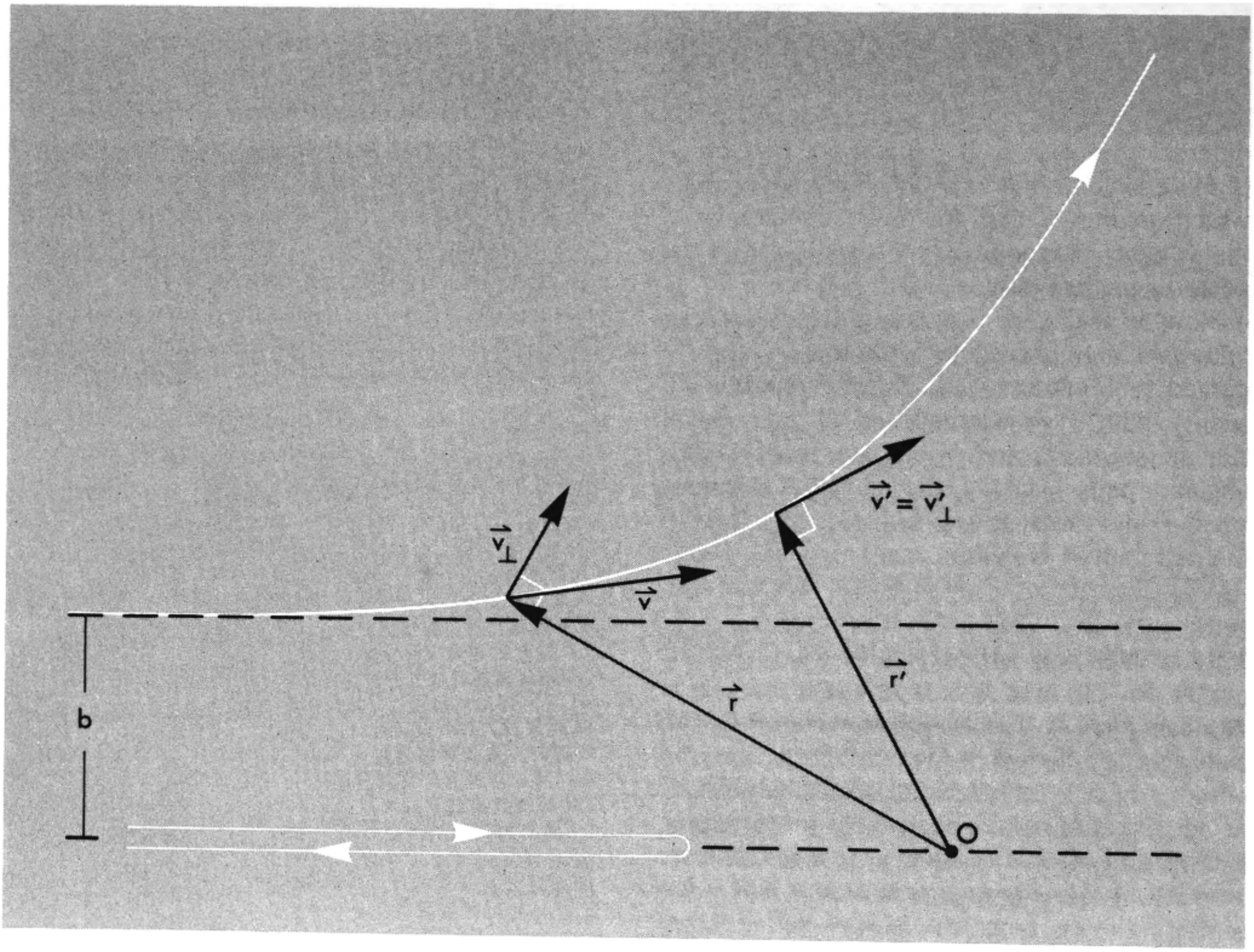
the total angular momentum remains constant also for interacting bodies, while the sum of the individual $rv_\perp$ does not.

In Fig. 19 – 4, notice that the product $rv_\perp$ equals the product $r_\perp v$, since either is twice the area of the triangle *OAB*. We can also write the same product simply as $rv \sin \theta$. We shall make use of that and write $L = mrv_\perp = mr_\perp v = mrv \sin \theta$, according to our convenience. Notice that the definition of angular momentum contains the position vector $\vec{r}$, which is measured from a given point. Thus even a single body moving with constant velocity has generally different values of angular momentum, depending on the point to which it is referred.

19–3 ENERGY, ANGULAR MOMENTUM, AND TRAJECTORIES

The total energy (kinetic plus potential) of a body moving in a central field of force remains constant all along the trajectory. So does the angular momentum of the body with respect to the center of force. For different values of the energy and angular momentum we should expect different trajectories.

Consider, for example, the motion of an alpha particle (the nucleus of a helium atom) in the field of a gold nucleus. Near the nucleus the alpha particle moves in a repulsive central force field. As it comes closer to the gold nucleus its kinetic energy decreases and its potential energy increases. How close can the alpha particle get to the nucleus? The conservation of energy tells us that it can at best get so close to the nu-



19–5 Two trajectories (shown in white) of an alpha particle near a nucleus located at O. The initial velocity is the same in both cases, but the distance of closest approach is smaller in the lower trajectory.

cleus that all its initial kinetic energy (which it had far away from the nucleus) is converted into potential energy:

$$\frac{1}{2}mv^2_{\text{initial}} = U.$$

At that distance the alpha particle will be at rest; it has, therefore, no angular momentum. The conservation of angular momentum then tells us that the alpha particle had no angular momentum to begin with: that is, $mrv_\perp = 0$. Since r is not zero, then $v_\perp = 0$, and the alpha particle was moving straight toward the gold nucleus (Fig. 19 – 5). If, on the other hand, the alpha particle (starting with the same velocity) is not initially moving exactly toward the gold nucleus, then it must have some angular momentum. If its aim is off by the distance b, the angular momentum of the alpha particle around the nucleus is

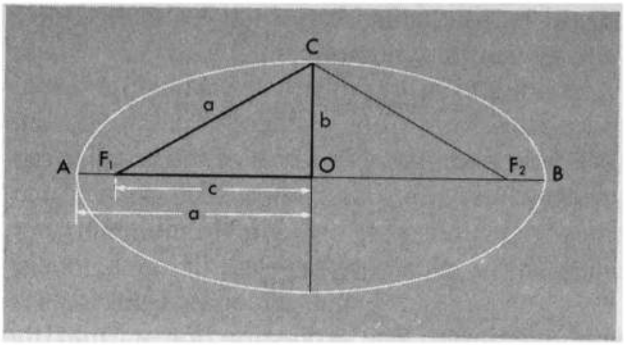
$$L = mrv_\perp = mvr_\perp = mv_{\text{initial}}b$$

We see immediately that the alpha particle will not get as close to the nucleus as before, because at the distance of closest approach r' there will be a velocity $\vec{v}_\perp'$ such that $mr'v_\perp' = mv_{\text{initial}}b$. Consequently there is always some kinetic energy, and for a repulsive force this means that the alpha particle never gets as close as it does when it has no angular momentum (Fig. 19 – 5.) We see that two alpha particles approaching the same nucleus in the same direction with the same kinetic energy will have different trajectories if they have different angular momenta with respect to the nucleus.

19–4 A SPECIAL CASE: THE MOTION OF SATELLITES

In the previous section we discussed qualitatively the relation between energy, angular momentum, and the trajectory of an object moving in a central field of force. For the motion of satellites we can study this relation in detail. In our calculations we shall use only the inverse-square nature of the gravitational force; our conclusion will be valid for any other inverse-square force, in particular for electric forces.

As we know, earth satellites move in elliptical orbits with the earth at one focus. An ellipse is defined as the locus of points P such that the sum of the distances from P to each of two fixed points (known as foci) is a constant $2a$ (Fig. 19 – 6). Each ellipse may be characterized by this constant length and by the separation $2c$ of the foci. An ellipse can also be described in terms of the semi-major axis a and the semi-minor axis b. The relation between the two descriptions is shown in Fig. 19 – 6.

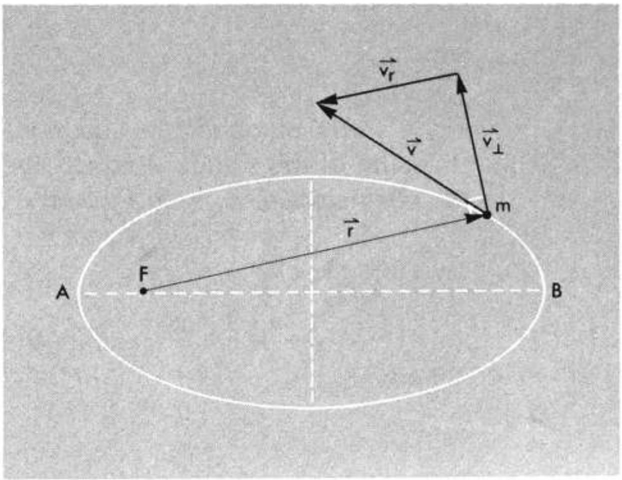


19–6 Since A lies on the ellipse, $AF_1 + AF_2 = 2a$. By symmetry, $AF_1 = F_2B$, so $AB = 2a$ and the semi-major axis equals a. Since C also lies on the ellipse, $CF_1 + CF_2 = 2a$. Again by symmetry, $CF_1 = CF_2$ so $CF_1 = a$. And since OCF_1 is a right triangle, $b^2 = a^2 - c^2$.

How are a and b related to the total energy E and the angular momentum L of the satellite? First let us find an equation which will relate E and L.

We resolve the satellite's velocity $\vec{v}$ into two components, one in the direction of the position vector and one perpendicular to it:

$$\vec{v} = \vec{v}_r + \vec{v}_\perp$$



19–7 The velocity $\vec{v}$ is split into a component $\vec{v}_r$ parallel to $\vec{r}$ and a component $\vec{v}_\perp$ perpendicular to $\vec{r}$.

(Fig. 19 – 7). Hence $v^2 = v_r^2 + v_\perp^2$, and the kinetic energy can be written

$$E_K = \tfrac{1}{2}mv^2 = \tfrac{1}{2}mv_r^2 + \tfrac{1}{2}mv_\perp^2$$

We now express $v_\perp$ in terms of the angular momentum L:

$$v_\perp = \frac{L}{mr}$$

and we insert the expression for $v_\perp$ into the equation for E_K:

$$E_K = \frac{1}{2}mv_r^2 + \frac{1}{2}\frac{L^2}{mr^2}$$

The total energy becomes then

$$E = E_K + U_r = \frac{1}{2}mv_r^2 + \frac{1}{2}\frac{L^2}{mr^2} - G\frac{Mm}{r}$$

where M and m are the masses of the earth and of the satellite respectively and G is the universal gravitational constant. This expression for the energy holds for all points on the orbit. At the farthest and closest points from the earth $v_r = 0$. We shall denote the distances to these points by r_0. Then at these points the expression for the energy is reduced to

$$E = \frac{L^2}{2mr_0^2} - G\frac{Mm}{r_0}$$

Multiplying both sides by $\frac{r_0^2}{E}$, we find that this is a quadratic equation for r_0:

$$r_0^2 + \frac{GMm}{E}r_0 - \frac{L^2}{2mE} = 0$$

We know from Fig. 19 – 6 that the two solutions of this equation are

$$r_0 = a + c \quad \text{and} \quad r_0 = a - c$$

Furthermore, the study of quadratic equations tells us that the sum of the solutions equals the negative of the coefficient of r_0; in this case

$$(a + c) + (a - c) = -\frac{GMm}{E}$$

Also, the product of the solutions equals the term not containing r_0:

$$(a + c)(a - c) = \frac{-L^2}{2mE}$$

From the first equation we have

$$2a = -\frac{GMm}{E} \quad \text{or} \quad a = -\frac{GMm}{2E}$$

From the second equation; using the relation between a, c, and the semi-minor axis b (see Fig. 19 – 6), we have

$$a^2 - c^2 = b^2 = \frac{-L^2}{2mE}$$

or

$$b = \frac{L}{\sqrt{-2mE}}$$

(Since the satellite is bound, the total energy E is negative. Thus we are taking the square root of a positive quantity. See Section 18 – 5.)

These equations show how the shape of the elliptical orbit depends on the energy and angular momentum of the satellite. Notice that the semi-major axis a of the ellipse depends only on the energy. If we launch a satellite, the major axis of its elliptical orbit will depend only on where it separates from its carrier and on its speed at the moment. It does not depend on the direction of its motion.

On the other hand, the semi-minor axis will depend on the direction of motion at the instant of separation. Since $L = mrv \sin \theta$, and b is proportional to L for a fixed energy, we see that the largest b will occur when the velocity is at right angles to the position vector.

The distance of closest approach to the earth is

$$r_{\min} = a - c = a - \sqrt{a^2 - b^2}$$

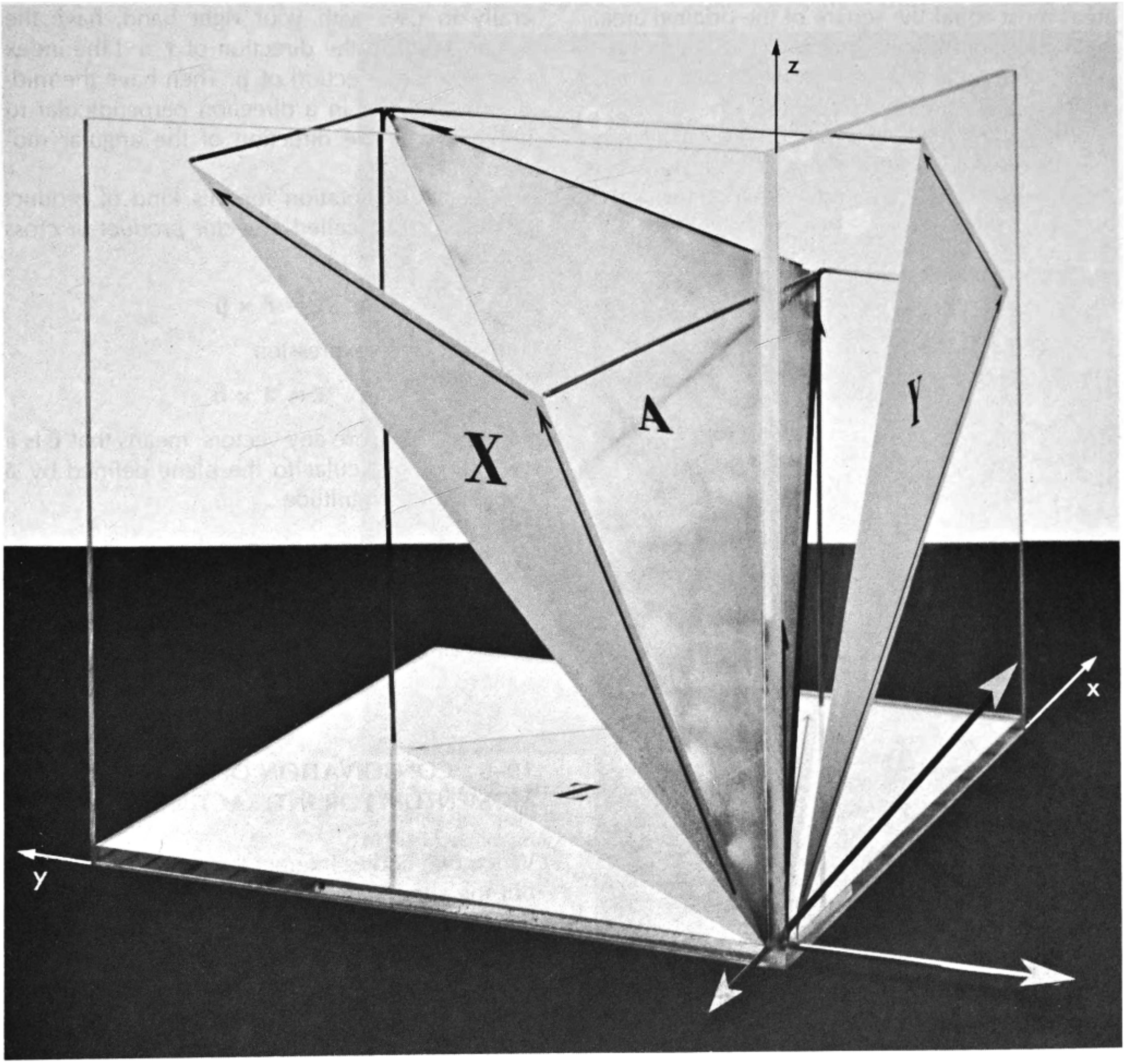
If b is small, the square root term is close to a and hence $r_{\min}$ is small. In this case, the ellipse is very flat with its foci very close to the ends of the major axis, The largest value b can have, for a suitable value of the energy, is a. Then $r_{\min} = a$, and the ellipse is reduced to a circle, its two foci coinciding with the center of the circle. Since b is proportional to L, we see that for a given energy the largest angular momentum is obtained in a circular orbit.

19–5 ANGULAR MOMENTUM: A VECTOR?

We have learned in the preceding section that the shape of the elliptical orbit of a satellite is determined by the satellite's energy and the angular momentum. But the shape alone does not specify the orbit completely; the ellipse lies in a plane, but the energy and the quantity $rp_\perp$ do not tell us how this plane is oriented in space. Must we look for additional information in order to specify the orientation of the plane of the orbit in space, or have we simply not used all the information at our disposal? Can you see that it is the latter which is true?

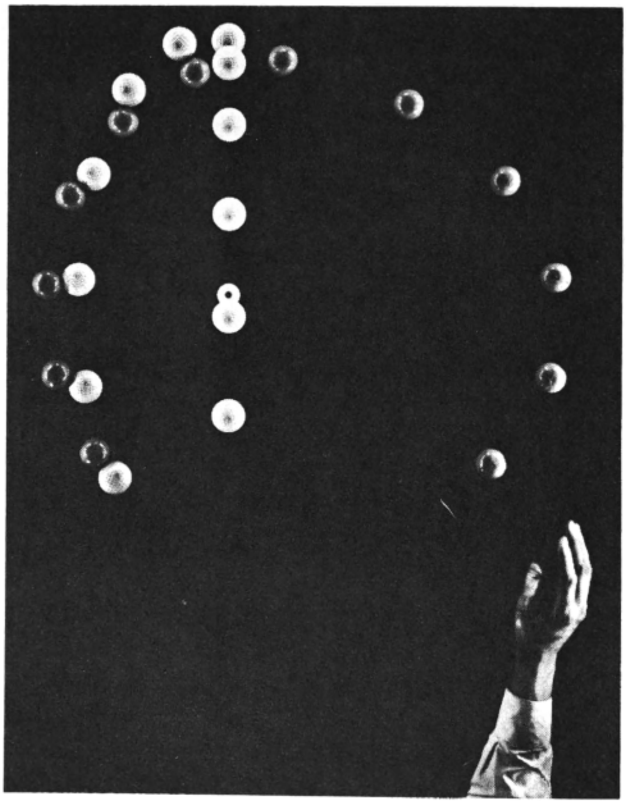
The position vector $\vec{\mathbf{r}}$ and the momentum vector $\vec{\mathbf{p}}$ at any time define a plane. This plane contains the whole orbit because the force which changes $\vec{\mathbf{p}}$ is always along $\vec{\mathbf{r}}$, so the momentum never acquires a component outside of the plane. We could, therefore, describe the orbit in more detail by stating the value of the energy, the angular momentum, and the orientation in space of the plane containing the orbit. We can describe the orientation uniquely by giving the direction of any line perpendicular to the plane, since all lines perpendicular to the plane are parallel to one another.

The direction of such a line is perpendicular to both $\vec{\mathbf{r}}$ and $\vec{\mathbf{p}}$. Furthermore, $rp_\perp$ is the mag-



19–8 The triangular area X is the projection of the triangular area A on the y-z plane, Y is its projection on the x-z plane, and Z is its projection on the x-y plane. Measurements show that $X^2 + Y^2 + Z^2 = A^2$.

nitude of the angular momentum. It is tempting to ask whether this direction and magnitude combine to form a vector. We have learned that vectors are things that add and subtract like displacements. In particular we must be able to resolve any vector into vector components pointing along the axes of a rectangular coordinate system. Can we do this with a quantity which has the magnitude $rp_\perp$ and is perpendicular to both $\vec{r}$ and $\vec{p}$? You will recall that $rp_\perp$ is $2m$ times the area covered by the position vector per unit time. If such an area has the properties of a vector, its projections on three mutually perpendicular planes must add like vectors; the sum of the squares of the projected areas must equal the square of the original area. Figure 19 – 8 shows a model of a triangle projected onto three mutually perpendicular planes. Actual measurements show that the areas of the three projected triangles do indeed follow the rule of vector addition: that is, $X^2 + Y^2 + Z^2 = A^2$, where X, Y, Z are the areas of the three projected triangles and A is the area of the original triangle. Other triangles with different orientation give the same relationship. We chose to prove the vector nature of angular momentum in an empirical way in order to give you a better feeling for this vector.



19–9 A white golf ball and a dark putty ball of equal mass are both suspended from a high ceiling. The point of attachment is directly above the black dot in the middle of the photograph. Before the collision, the golf ball has no angular momentum with respect to the black dot.

One more point needs clarification. The vectors $\vec{r}$ and $\vec{p}$ define a plane which can be characterized by a line perpendicular to it. But a vector has a definite direction. Which of the two opposite directions along this line is the direction of the angular momentum vector? It can be either. The answer is a matter of choice, provided we are consistent and always choose it the same way. The following procedure is generally in use: with your right hand, have the thumb point in the direction of $\vec{r}$ and the index finger in the direction of $\vec{p}$. Then have the middle finger point in a direction perpendicular to both. This is the direction of the angular momentum.

A standard notation for this kind of product of two vectors, called a *vector product* or *cross product,* is

$$\vec{L} = \vec{r} \times \vec{p}$$

In general, the expression

$$\vec{c} = \vec{a} \times \vec{b}$$

where $\vec{a}$ and $\vec{b}$ are any vectors, means that $\vec{c}$ is a vector perpendicular to the plane defined by $\vec{a}$ and $\vec{b}$, with magnitude

$$c = ab_\perp = a_\perp b$$

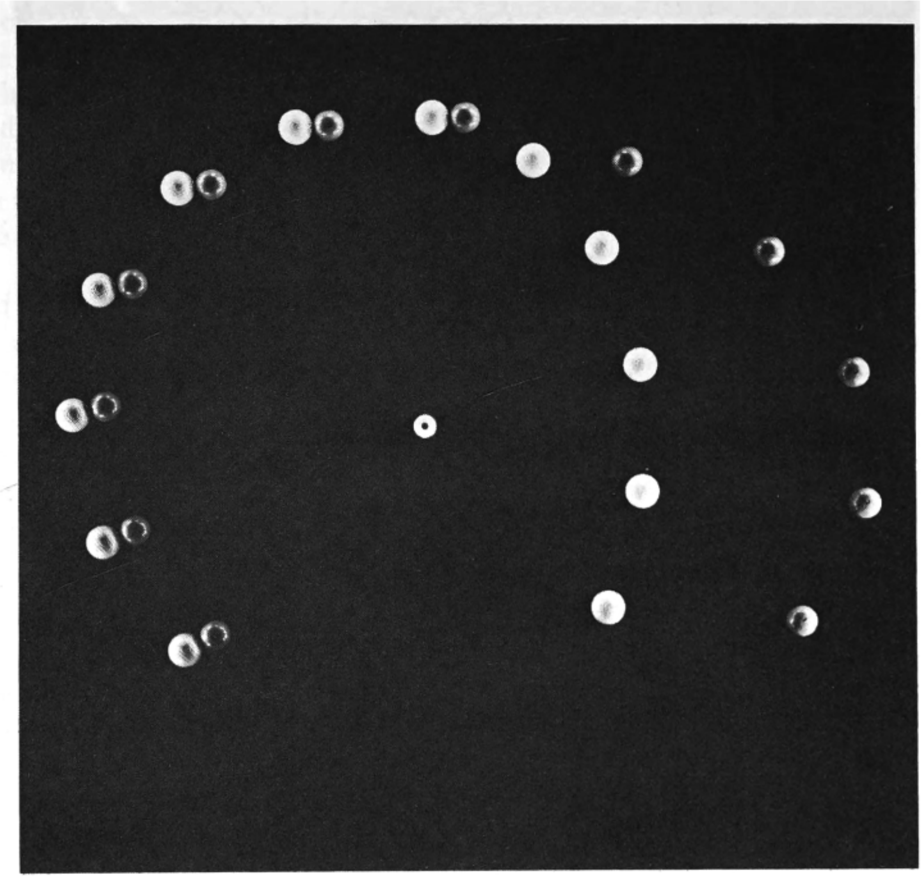
Using the right-hand rule, if your thumb points in the direction of $\vec{a}$ and your index finger points in the direction of $\vec{b}$, then the middle finger of your right hand indicates the direction of $\vec{c}$.

19–6 CONSERVATION OF ANGULAR MOMENTUM FOR INTERACTING BODIES

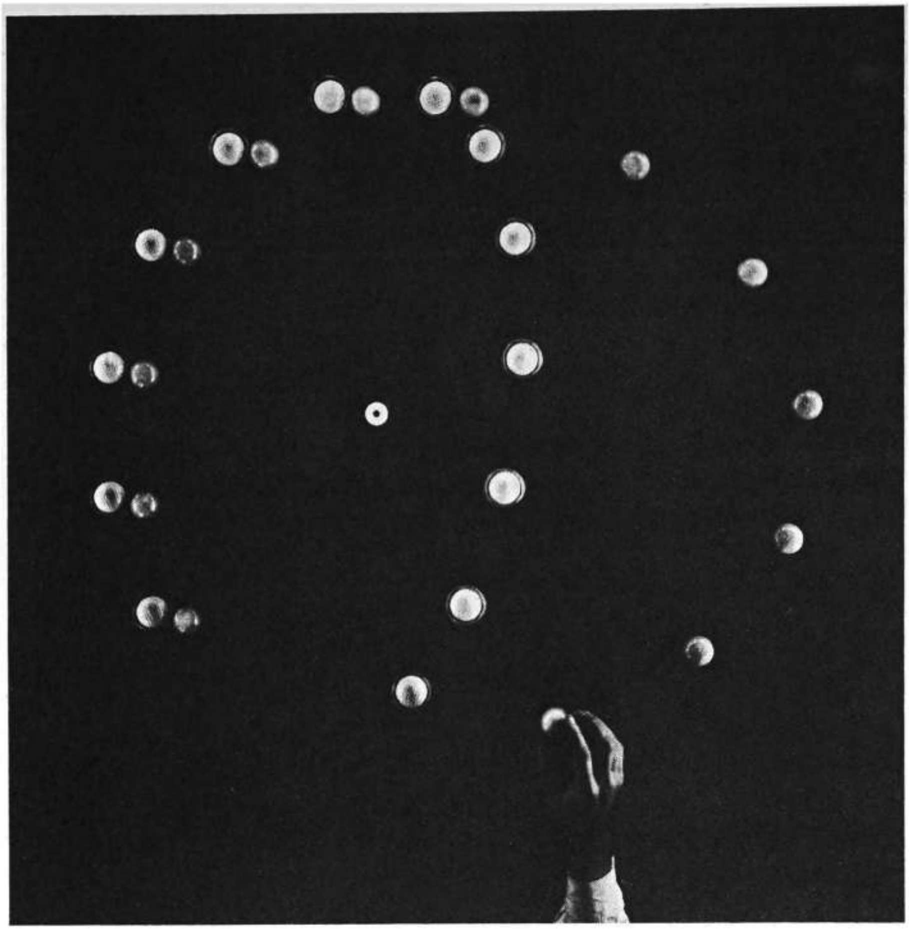
When two bodies interact their momenta change but the sum of their momenta remains constant if no external forces are acting. Let us see whether a similar situation prevails for the angular momentum when two bodies interact. Figure 19 – 9 shows two orbiting objects, a golf ball and a putty ball of equal mass, which collide and stick together. The position vector of the putty ball initially sweeps out an area at a constant rate. The golf ball initially orbits right through the center of force (marked with a black dot on the white circle) so that its position vector sweeps out zero area per unit time. The golf ball has no angular momentum before the collision, with respect to the force center. After the collision the two balls rotate about each other to a certain extent, but this makes only small changes in the position of each object. We can neglect this contribution to the area in these experiments and use the position of the center of mass of the two objects in our measurements. Careful measurements show that the position vector from the center of force to the center of mass of the combined balls after the collision sweeps out almost exactly half the area that was swept out by the position vector from force center to putty ball before the collision. But now it is double the mass that is moving, so that the angular momentum remains unchanged.

In Fig. 19 – 10 we have a similar system except that both balls sweep out sizable areas per unit time before the collision. How do the areas swept out before the collision compare with the area swept out jointly after the collision? What is the total angular momentum after the collision?

Let us now analyze a collision of two objects



19–10 A similar arrangement to that in Fig. 19–9, except that here both balls have angular momentum with respect to the black dot prior to the collision.



19–11 The golf ball is loaded with a belt of putty, so that its mass is 1.50 times that of the putty ball.

of different mass. Figure 19 – 11 shows a collision like those described above, except that the golf ball is loaded with a belt of putty so that it has a mass 1.50 times the mass of the solid putty ball. The data are given in Table 1. The angular momentum before the collision is $(1.50)(2.51) + (1.00)(7.90) = 11.7$. After the collision it is $(1.50 + 1.00)(4.65) = 11.6$. We see that the sum of the angular momenta is the same before and after the collision. Notice that both balls move in the same plane and rotate in the same sense. Hence their angular momentum vectors are parallel.

Many other experiments with elastic or partially elastic collisions, including situations where the colliding objects move in different planes, yield similar results; the vector sum of the angular momenta of the colliding bodies is conserved. (See Section 19 – 10.)

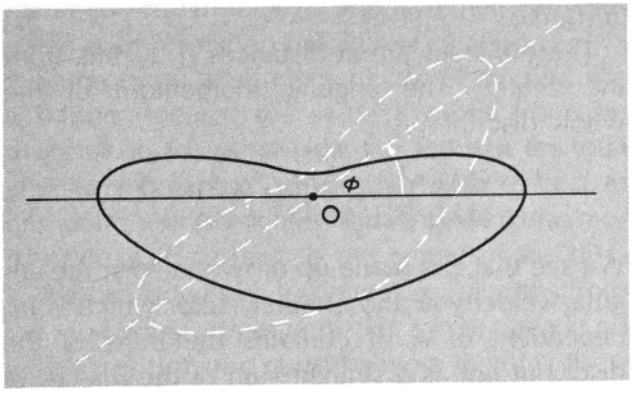
TABLE 1

BEFORE COLLISION						AFTER COLLISION		
GOLF BALL			PUTTY BALL			COMBINED BALLS		
r	$v_\perp$	$rv_\perp$	r	$v_\perp$	$rv_\perp$	r	$v_\perp$	$rv_\perp$
3.31	0.76	2.52	4.24	1.87	7.93	3.81	1.21	4.61
2.52	0.99	2.49	4.78	1.66	7.93	3.75	1.25	4.69
1.78	1.40	2.49	5.13	1.53	7.85	3.50	1.32	4.62
1.87	1.34	2.51	5.13	1.53	7.85	3.22	1.45	4.67
2.62	0.96	2.52	4.77	1.66	7.92	3.07	1.51	4.64
$rv_{\perp av} = 2.51$			$rv_{\perp av} = 7.90$			$rv_{\perp av} = 4.65$		

19–7 THE ROTATION OF RIGID BODIES

So far we have discussed only motions of bodies along trajectories which are large compared to the size of the bodies themselves. This limitation justified our ignoring the small differences in the trajectories of the different parts of the body. (See Fig. 19 – 1.) If the size of the moving body is comparable to the size of the orbit, we can no longer ignore these differences, and the description of the motion becomes, in general, more complicated. In this section we shall go to the other extreme and discuss the rotation of a body around a fixed axis passing through the body. The points along the axis of rotation do not move at all, while those off the axis move faster the larger their distance from it. Here, naturally, we cannot ignore the difference in motion of the various parts of the body; nevertheless, the description of the motion becomes simple if we limit the discussion to bodies whose shape remains, for all practical purposes, unchanged — rigid bodies.

Consider a block of wood which rotates around a fixed axis. Different little chunks of wood in the block move with different speeds depending on their distance from the axis. Therefore, the individual speeds will not give us a convenient way of describing the motion of the block as a whole. Now imagine that we draw a straight line which is fixed in the block and passes through the axis of rotation at right angles (Fig. 19 – 12). As the block rotates, the line rotates with it. The amount of rotation of the line is expressed in terms of the angle between the original direction and its direction at a later time. If the block is rigid the position of any of the little chunks which make up the block, relative to the straight line, does not change. Every little chunk rotates through the same angle (ϕ in Fig. 19 – 12) and the angle of rotation describes the motion of the whole block.



19–12 A rigid body rotating through an angle ϕ about a fixed axis through O which is perpendicular to the plane of the paper.

As the block rotates the angle ϕ changes. Its average rate of change during a time Δt is the average angular velocity:

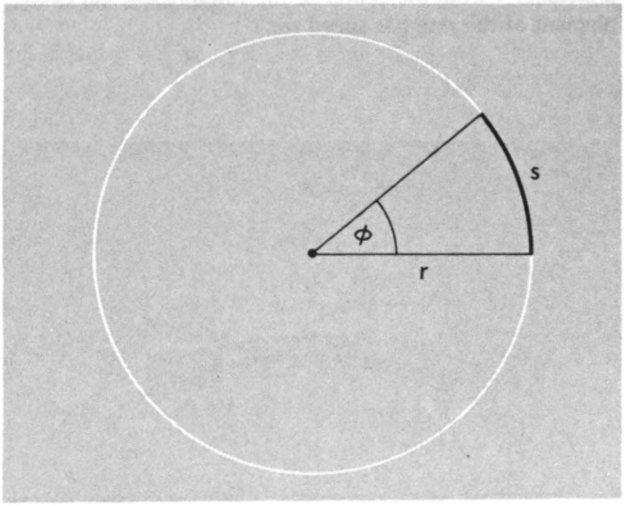
$$\omega_{\text{av}} = \frac{\Delta\phi}{\Delta t}$$

As with linear velocity, we can define an instantaneous angular velocity by the relation

$$\omega = \lim_{\Delta t\to 0} \frac{\Delta\phi}{\Delta t}$$

It is convenient to measure the angle of rotation in radians, that is, to measure ϕ as the length of a circular arc s divided by the radius r (Fig. 19 – 13). Then the length of the arc traveled by a point a distance r from the axis is $s = r\phi$ and its speed is $v = r\omega$.

The angular velocity, like the angle of rotation, applies to the whole block. The situation would not be so simple if instead of a block of wood we had a block of jelly; various chunks in the block of jelly would have not only different speeds but different angular velocities, because the jelly might twist and change its shape during the rotation. It would not be a rigid body.



19–13 Angle ϕ in radians is arc length s divided by radius r: $\phi = s/r$ or $s = r\phi$. If ϕ changes with time, s changes also, but r remains unchanged. Therefore

$$v = \frac{\Delta s}{\Delta t} = \frac{r\Delta\phi}{\Delta t} = r\omega$$

19–8 THE ANGULAR MOMENTUM OF RIGID BODIES; MOMENT OF INERTIA

We think of a rigid body as being made up of many small chunks of matter which keep their relative positions. The total angular momentum of the body is the sum of the angular momenta of all these chunks, often called "mass elements." Figure 19 – 14 shows a thin, uniform disc rotating around a perpendicular axis through its center. What is the relation between its angular momentum with respect to its center and its angular velocity? Let us concentrate first on one mass element Δm at a distance r from the center. If the disc rotates with angular velocity ω in the sense shown in the figure, the mass element has a velocity at right angles to $\vec{r}$ of magnitude $v = r\omega$. Following the "right hand" convention, the angular momentum $\Delta\vec{L}$ of the mass element under consideration has the magnitude

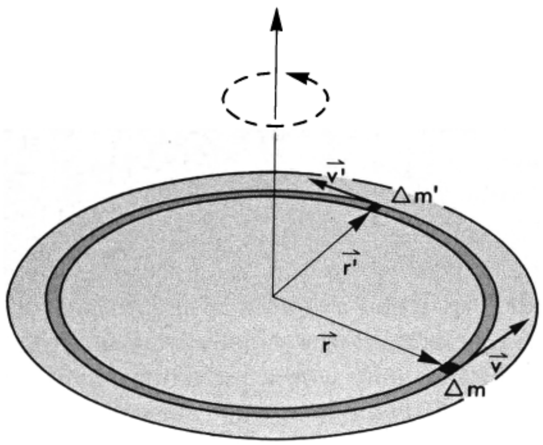
$$\Delta L = \Delta mrv = \Delta mr^2\omega$$

and points up along the axis of rotation.

Now let us add another mass element $\Delta m'$ at the same distance r as Δm. The magnitude of its angular momentum is $\Delta L' = \Delta m' r^2\omega$, and since it rotates in the plane and in the same sense as Δm, the direction of its angular momentum is also up the axis of rotation. The sum of the angular momenta of these two mass elements is

$$\begin{aligned}\Delta L + \Delta L' &= \Delta mr^2\omega + \Delta m'r^2\omega \\ &= (\Delta m + \Delta m')r^2\omega\end{aligned}$$

We can continue to add mass elements to mass elements at the same distance until we have covered the whole ring (Fig. 19 – 14). The angular momentum of the thin ring is then $m_{\text{ring}}r^2\omega$. This is the same as if the mass were all concentrated in one lump at a distance r and rotating with the same angular velocity ω.



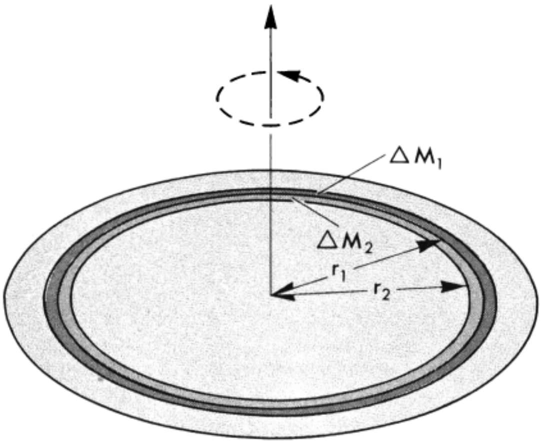
19–14 The shaded ring is made up of all mass elements a distance r from the center of the rotating disc, so every element of the ring has speed rω.

We can now find the angular momentum of the whole disc by dividing it into narrow rings (Fig. 19 – 15) and summing up the angular momenta of all the rings. The only factor in the angular momenta of the individual rings which is common to all of them is the angular velocity. Their radii and masses are different. To distinguish the masses of the rings from those of the individual chunks, we denote them by ΔM_1, ΔM_2, etc.

These masses are at distances r_1, r_2, etc. from the center. The angular momentum of the whole disc then is

$$\begin{aligned}L &= (\Delta M_1 r_1{}^2\omega + \Delta M_2 r_2{}^2\omega + \ldots) \\ &= (\Delta M_1 r_1{}^2 + \Delta M_2 r_2{}^2 + \ldots)\omega\end{aligned}$$

We see that L is made up of two factors: the angular velocity ω and another factor which is independent of ω. It contains the mass of the disc, but not as a straight sum of the masses of all the rings which make up the disc. The contribution of each ΔM increases with the square of its distance from the center of the disc.



19–15 The ring of mass ΔM_1 and the ring of mass ΔM_2 have the same angular velocity but different radii.

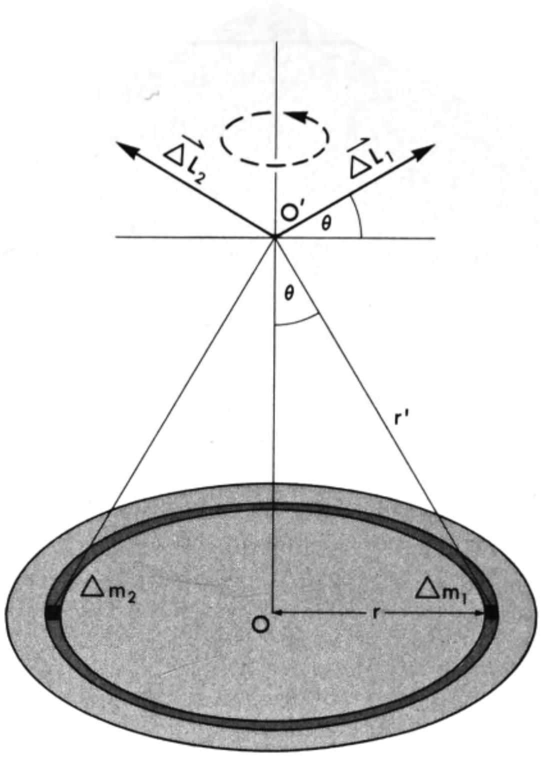
We have calculated the angular momentum of a thin disc with respect to its center. What would the angular momentum be with respect to a point on the axis but outside the disc (O' in Fig. 19 – 16)? Consider first the mass element Δm_1 at the right-hand side of the figure, a distance r from the axis. The position vector of this mass element for O' has the length $r' = \frac{r}{\sin\theta}$. Its velocity vector points into the plane of the paper and has the magnitude $v = r\omega$ as before. The angular momentum of the mass element has, therefore, the magnitude $\Delta L = \Delta m r^2 \omega / \sin\theta$ and a direction perpendicular to r' in the plane of the paper as shown in the figure. Because of symmetry, the angular momentum of the mass element on the left will have the same magnitude and the same inclination toward the axis of rotation as the element on the right. We see that the vector components of the two angular momenta parallel to the disc cancel, whereas the components parallel to the axis of rotation add up (Fig. 19 – 17). The vertical components have the magnitude $\Delta L \sin\theta$. Substituting for ΔL the expression which we have just calculated, we find

$$\Delta L \sin\theta = \Delta m \frac{r^2\omega}{\sin\theta} \sin\theta = \Delta m r^2 \omega$$

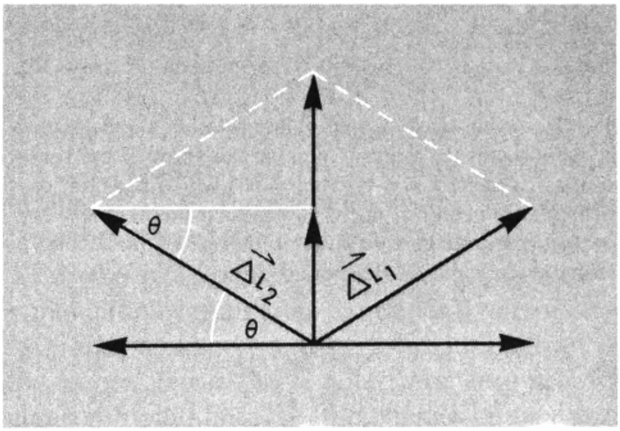
The vector sum of the angular momenta of the two equal and symmetrically located mass elements with respect to the point O' is equal to the sum of the angular momenta of the same mass elements with respect to O. Since the point O' was picked arbitrarily, our conclusion holds for any point along the axis. Furthermore, our argument will hold for any two mass elements equal in size and symmetrically located with respect to the axis of rotation.

Since every mass element Δm of the disc has a partner located the same distance from the axis but on the other side, we see that the total angular momentum of the disc with respect to any point along this axis equals the angular momentum with respect to the center of the disc.

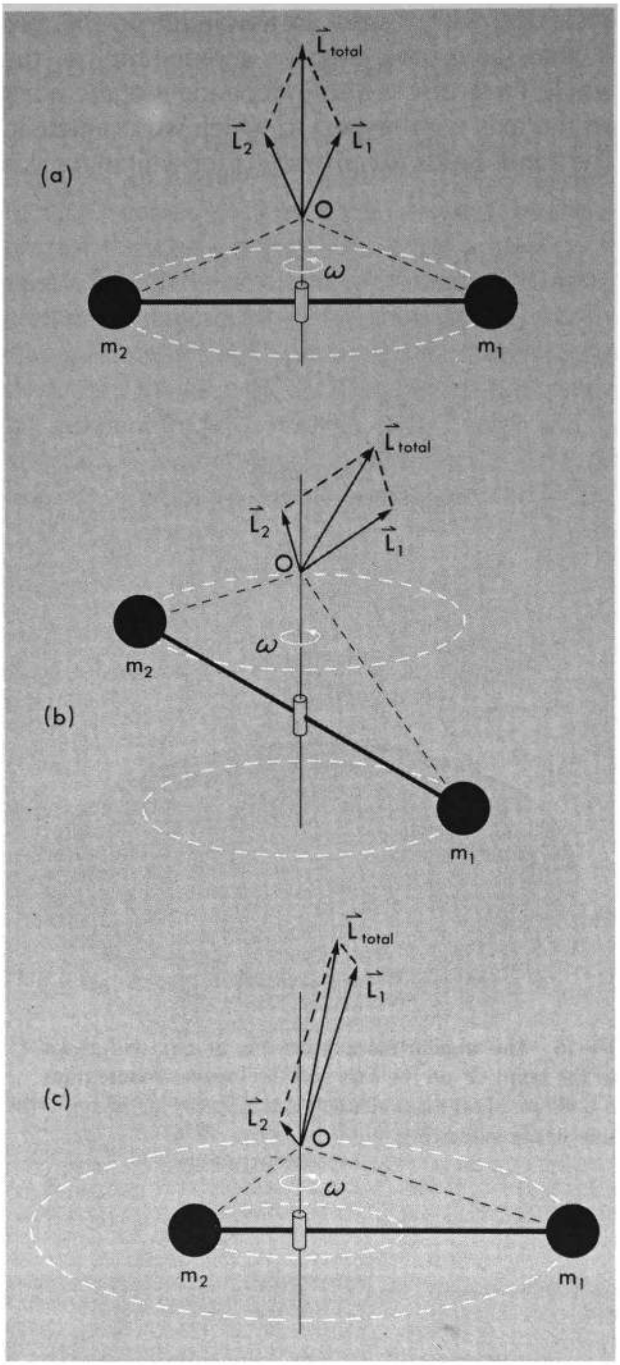
We can draw two important conclusions from the foregoing argument. First, we need no longer restrict our consideration to thin discs. We can always imagine a thick disc divided into many thin discs, and since for rotation about the symmetry axis the angular momentum of each one with respect to any point on the axis is the same, the angular momentum of the whole thick disc is also independent of the point on the axis with respect to which we evaluate it. The same holds for any object for which the axis



19–16 The angular momentum $\overrightarrow{\Delta L}_1$ of Δm_1 with respect to the point O′ on the axis and the angular momentum $\overrightarrow{\Delta L}_2$ of an equal mass element Δm_2 located at the opposite side of the same ring.



19–17 The horizontal vector components of $\overrightarrow{\Delta L}_1$ and $\overrightarrow{\Delta L}_2$ from Fig. 19–16 cancel, while the vertical components add.



19–18 A dumbbell-shaped object rotating about (a) a symmetry axis (b) a non-symmetry axis through the center of mass, and (c) a non-symmetry axis which does not pass through the center of mass. The angular momentum vectors in each figure relative to the point O lie in the plane of the paper at the instant shown.

of rotation is an axis of symmetry, such as a uniform rectangular block with the axis of rotation perpendicular to one of the surfaces and passing through the center of the block. For such objects, the angular momentum points along the axis of rotation and has a magnitude

$$L = (\Delta M_1 r_1^2 + \Delta M_2 r_2^2 + \dots)\omega$$

where the expression in parentheses, the sum of the products of the mass elements times the squares of the distances from the axis of rotation, is called the *moment of inertia,* and is denoted by the letter J. Thus $L = J\omega$. The moment of inertia of any object depends on its mass, its shape, and the location of the axis of rotation. It can be determined experimentally and—for not-too-complicated shapes—also calculated. For the moment of inertia of the disc with respect to a perpendicular axis through its center the result is $J_{\text{disc}} = \frac{1}{2}MR^2$ where M is the mass and R is the radius of the disc. A ring of mass and radius equal to that of the disc has a moment of inertia twice as big: $J_{\text{ring}} = MR^2$. If both are rotating with the same angular velocity, the ring has twice the angular momentum.

Our second conclusion is that if the axis of rotation is not a symmetry axis, at least some mass elements on one side of the axis of rotation will not have corresponding mass elements on the other side, and therefore the angular momentum does not in general point along the axis of rotation. Furthermore, its magnitude and direction may depend on the point on the axis about which it is taken. For example, consider a dumbbell, two point masses connected by a thin rigid rod, rotating about various axes, as in Fig. 19–18. The rotation in (a) is about a symmetry axis; in (b), about a nonsymmetry axis which passes through the center of mass of the dumbbell; and in (c), about a nonsymmetry axis which does not pass through the center of mass. The angular momentum shown for each mass element and for the entire object has been calculated with respect to an arbitrary point O on each axis at the instant the dumbbell lies in the plane of the paper.

As each dumbbell rotates with constant angular velocity, the total angular momentum in each of the three cases remains constant in magnitude, but in both (b) and (c), its direction changes continually. The tip of the angular momentum vector moves along a circle centered on the axis of rotation. It never points along the axis of rotation. For these rotations about nonsymmetry axes, the angular momentum is not equal to $(M_1r_1^2 + M_2r_2^2)\omega$. This product yields only the component of the angular momentum along the axis of rotation. (See Question 32.) These conclusions hold for any object rotating about a nonsymmetry axis.

19–9 TORQUE—THE RATE OF CHANGE OF ANGULAR MOMENTUM

So far we have studied primarily situations where the angular momentum remained constant. What determines its rate of change when it does not stay constant? We shall study first the simple case where the dimensions of the trajectory are large compared with the dimensions of the moving body, thus ignoring the differences in the motion of its parts. Furthermore, we shall treat an idealized motion where the forces act only for very short times at regular intervals, as we did at the beginning of this chapter.

In the fixed unit of time the body moves from A to B (Fig. 19–19). The area swept out by the position vector is OAB. At B a force $\vec{\mathbf{F}}$ acts for any short Δt in the direction $\overrightarrow{BN}$ different from $\overrightarrow{BO}$, but in the plane of the paper. This causes a change in velocity $\Delta\vec{\mathbf{v}} = \vec{\mathbf{F}}\Delta t/m$, which is shown in the figure as the vector $\overrightarrow{BK}$. In the next unit of time it will, therefore, arrive at C, sweeping out the area OBC.

To find the point C which the body will reach with its new velocity during the second time interval, we add to vector $\overrightarrow{BK}$ a vector $\overrightarrow{KC}$ equal to vector $\overrightarrow{AB}$. We wish to relate the difference in the area, $OBC - OAB$, to $\vec{\mathbf{F}}\Delta t$. We already know from Section 19–1 that if the force were directed toward O, there would be no change in the areas. Let us therefore resolve $\vec{\mathbf{F}}$ into two vector components, $\vec{\mathbf{F}}_r$ along the line BO and $\vec{\mathbf{F}}_\perp$ perpendicular to it:

$$\vec{\mathbf{F}} = \vec{\mathbf{F}}_r + \vec{\mathbf{F}}_\perp$$

Similarly, we write

$$\Delta\vec{\mathbf{v}} = \Delta\vec{\mathbf{v}}_r + \Delta\vec{\mathbf{v}}_\perp$$

Newton's law of motion tells us that

$$\Delta\vec{\mathbf{v}}_r = \frac{\vec{\mathbf{F}}_r\Delta t}{m} \quad \text{and} \quad \Delta\vec{\mathbf{v}}_\perp = \frac{\vec{\mathbf{F}}_\perp\Delta t}{m}$$

If only $\vec{\mathbf{F}}_r$ were present, the body would move to C' in the second unit of time. We already know that $OBC' = OAB$; hence

$$OBC - OAB = OBC - OBC'$$

This difference can be written as

$$\Delta S = OBC - OBC' = \tfrac{1}{2}OB \times CC'$$

In this equation $OB = r$, the length of the position vector and $CC' = \Delta v_\perp = F_\perp\Delta t/m$. Hence, we can rewrite the expression for ΔS in the form

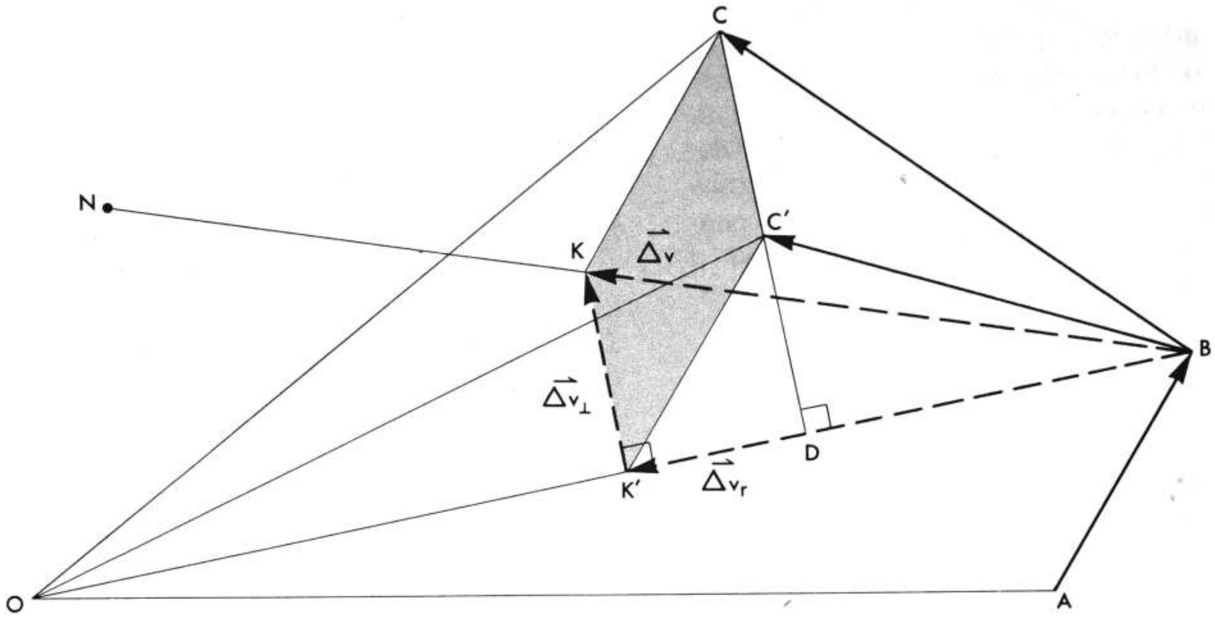
$$\Delta S = \frac{rF_\perp\Delta t}{2m}$$

Multiplying both sides by $2m/\Delta t$ we get on the left-hand side the rate of change of the angular momentum: $2m\Delta S/\Delta t = \Delta L/\Delta t$, and on the right-hand side the product $rF_\perp$, the position vector times the component of the force perpendicular to it. In short

$$\frac{\Delta L}{\Delta t} = rF_\perp$$

The product $rF_\perp$ is called the *torque* and will be denoted by T. Notice that like the angular momentum the torque is also the product of the magnitude of one vector times the perpendicular component of another vector. We can, therefore, express the torque also as $T = r_\perp F$, or $T = rF\sin\theta$ where θ is the angle between the two vectors. Furthermore, we suspect that the torque is also a vector whose magnitude is $rF_\perp$ and whose direction is perpendicular to both $\vec{\mathbf{r}}$ and $\vec{\mathbf{F}}$. The question now suggests itself whether the equation $\Delta L/\Delta t = T$ is really an equation between vectors, that is, whether $\Delta\vec{\mathbf{L}}/\Delta t = \vec{\mathbf{T}}$.

Let us look now at the special case described in Fig. 19–19. The velocity vectors before and after the impulse, as well as the force vector, are all in the plane of the paper. By the right-hand convention, the angular momentum vectors before and after the impulse point straight out of the paper. Since the angular momentum is increased by the impulse, the change in the angular momentum, $\Delta\vec{\mathbf{L}}$, also points up. For the vector equation to hold, we need only to apply the right-hand convention to the torque: If the



19–19 The vectors $\vec{AB}$ and $\vec{BC}$ represent the velocities of the body before and after it receives an impulse directed toward point N. Since $\vec{KC} = \vec{K'C'} = \vec{AB}$, the shaded area is a parallelogram, and $CC' = KK' = \Delta v_\perp$.

thumb of the right hand points along $\vec{r}$, and the index finger points along $\vec{F}$, then when the middle finger is perpendicular to both it will point along $\vec{T}$. Using the vector product notation introduced in Section 19 – 5, we have

$$\vec{T} = \vec{r} \times \vec{F}$$

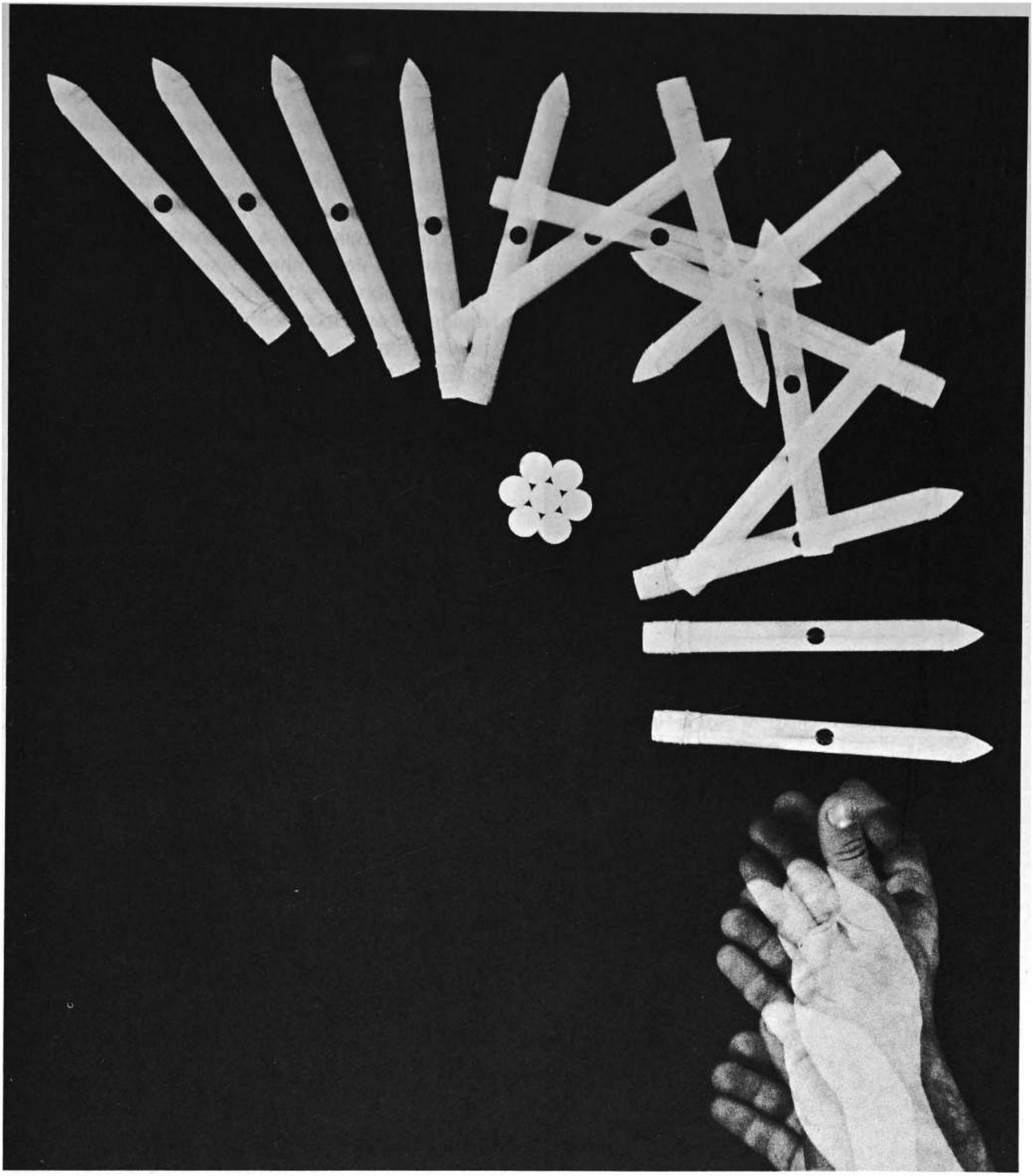
We can easily extend the derivation of the above equation to the more realistic one where the torque acts continuously, just as we did for the central force in Section 19 – 1, but this is left to the reader.

So far, we have discussed only the case where the torque changed the magnitude but not the direction of the angular momentum. You can convince yourself that the vector equation $\Delta\vec{L}/\Delta t = \vec{T}$ holds also when the torque is not parallel to the angular momentum and thus changes its direction. (See Question 40.)

19–10 ORBITAL ANGULAR MOMENTUM AND SPIN

In the preceding sections we discussed two extreme cases of motion: on the one hand was an object moving in an orbit large compared to its size, and on the other hand was a body rotating around a fixed axis. The angular momentum of the body in the first case has the magnitude $mrv_\perp$ and is perpendicular to both $\vec{r}$ and $\vec{v}$. In the second case, restricting ourselves to rotations around an axis of symmetry, we found the angular momentum to have the magnitude $J\omega$ and to point along the axis of rotation. Let us now use this knowledge to study the angular momentum of a body moving in an orbit which is comparable in size to the length of the body itself.

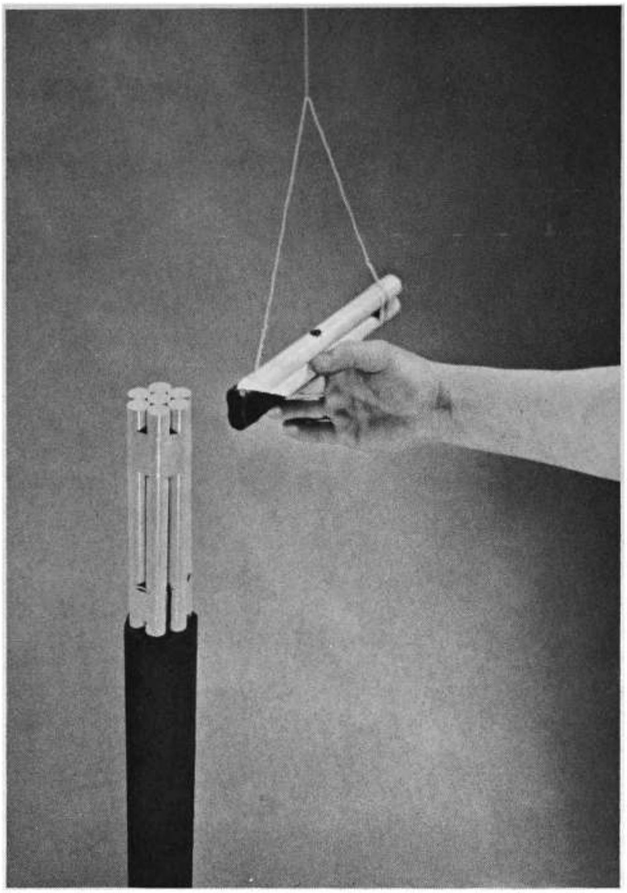
Figure 19 – 20 shows the motion of a rod. To distinguish between the two ends of the rod, one of them was covered with tape, thus giving the impression of a sharp end. In reality the rod is symmetrical; it is suspended at its center of mass, which is indicated by the black dot. The motion of the rod clearly can be described as being made up of two parts: the motion of the center of mass along the orbit and the rotation of the rod around an axis passing through the center of mass. This is not the only way of describing the motion of the rod. We can also, for example, separate the motion into the orbital motion of one end of the rod and a rotation around an axis passing through that end. Separating the motion of the center of mass has an important advantage: the angular momentum of the rod with respect to any point in space



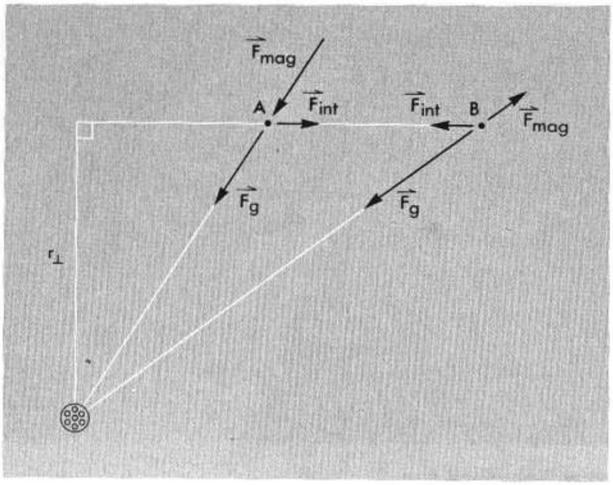
19–20 A suspended bar magnet moving in the field of a fixed magnet. The black dot on the bar magnet indicates the position of its center of mass. See details in Fig. 19–21.

separates into two additive parts, the angular momentum due to the motion of the center of mass and the angular momentum due to the rotation around the center of mass. The first part is commonly called the *orbital angular momentum* and the other the *spin angular momentum*, or just "spin" for short.

A formal proof of this separation is given in the box (page 360). Here we show it by a further study of Fig. 19 – 20. The rod in the



19–21 A close-up of the magnets used in Fig. 19–20.



19–22 The forces acting on an idealized magnet made of two "poles," A and B.

photograph is a bar magnet suspended from a high ceiling. In addition to gravity it is also acted upon by the magnetic forces of a magnet held fixed in a vertical position directly under the point of suspension. The upper end of the fixed magnet is in the same horizontal plane as the moving magnet (Fig. 19 – 21). As you know, the north poles or south poles of two magnets repel each other, whereas a north pole and a south pole attract each other. In our case we have to pay attention only to the forces exerted on the moving magnet by the upper pole of the fixed magnet, because the lower pole is too far away to have any considerable effect. Indeed, you can compare the length of the moving magnet in the various positions on the photograph and see that it is always the same, indicating that the magnet was not tilted with respect to the plane of the photograph.

There is one other force we have to consider: the internal force holding the parts of the magnet together. To simplify the picture, let us imagine for a moment that the magnet consists of two "pole" pieces *A* and *B* held together by an internal force. All the forces acting on these "poles" are shown schematically in Fig. 19 – 22. It is clear from the figure that while the gravitational and magnetic forces are central forces, the internal forces are not. The internal force acting on "pole" *A* exerts a torque $r_\perp F_A$ which equals the rate of change of the angular momentum of "pole" *A*. Similarly, the internal force acting on "pole" *B* exerts a torque $r_\perp F_B$ which equals the rate of change of the angular momentum of "pole" *B*. Now, the forces holding the pieces of a solid together are Newtonian forces —that is, they are equal and opposite and act along the line connecting the two parts. Hence $\vec{F}_A = -\vec{F}_B$ and the torques on *A* and *B* are also equal and opposite. Therefore the change of the angular momentum of "pole" *A* is minus the change of *B*, or in other words, the sum of the angular momenta of both is not changed by the internal forces. The central forces certainly do not change the angular momentum of either of the "poles"; we conclude that the total angular momentum of our simplified magnet is conserved. Our line of reasoning can be extended to any number of pieces, and therefore to the real magnet.

In short, we predict that the angular momentum of the magnet shown in Fig. 19 – 20 is conserved throughout the motion. Let us examine the figure in detail. We choose the mass of the magnet as the unit of mass and measure $rv_\perp$ for the center-of-mass motion. This will give us the orbital angular momentum. The results are shown in the first column of Table 2. The orbital angular momentum is not conserved. What about the spin?

To find the spin we have measured the angular velocity (in radians per unit time) and multiplied it by the moment of inertia. (See Section 19 – 8.) The moment of inertia of a rod of length L with respect to an axis through its center is $mL^2/12$. Since we chose the mass of the magnet as unity, we need only to multiply the angular velocity by $L^2/12$. The results are shown in the second column of the table. The spin also varies during the motion. The third column shows the sum of the orbital angular momentum and the spin. It is constant to within the accuracy of the experiment. We thus have illustrated the contention made earlier in this section that the total angular momentum of a body can be separated into the orbital angular momentum of the center of mass and the spin around the center of mass.

For macroscopic bodies the separation of orbital angular momentum and spin may seem to be just a matter of convenience. After all, we may say, the spin is nothing but ordinary angular momentum of chunks of matter moving in circles around an axis. Actually this separation goes much further. The study of atomic particles such as electrons and protons reveals that they too possess both orbital angular momentum and spin. They behave like tiny gyroscopes. But, unlike the spin of the macroscopic bodies, their spin cannot be reduced to any kind of orbital motion around their "center of mass."

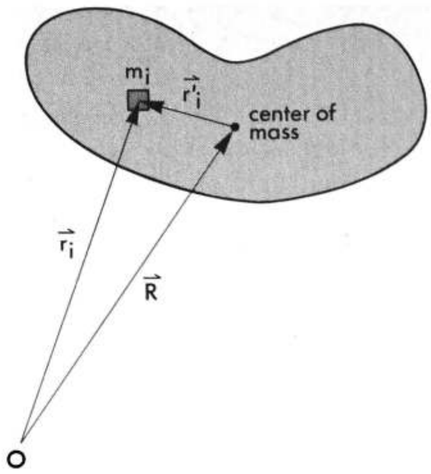
19–11 CONSERVATION OF ANGULAR MOMENTUM IN GENERAL

In the last section we have seen that the total angular momentum with respect to any point of two bodies, interacting only with each other via a Newtonian force, is conserved. But some forces in nature are not Newtonian. Most important among these are the forces responsible for the emission of light. In Section 16 – 7, we learned that in processes involving the emission or absorption of light, momentum is not conserved, if we overlook the momentum carried away by the light. The forces in question are not equal and opposite and they do not point along the line connecting the interacting bodies. It is not surprising, on the basis of the preceding section, that the angular momentum (orbital and spin) of the interacting bodies is not conserved when light is emitted. But it is reassuring to know that the light which is emitted or absorbed also carries angular momentum which restores the overall conservation.

The conservation of angular momentum when there is no external torque seems to be a fundamental law of nature, just like the conservation of linear momentum when there is no external force. In all cases where there appeared to be a violation of this law, it was found that something was overlooked, something which escaped detection and carried the missing angular momentum with it.

TABLE 2

ORBITAL ANGULAR MOMENTUM	SPIN	TOTAL ANGULAR MOMENTUM
6.1	0.2	6.3
5.7	0.6	6.3
4.9	1.3	6.2
4.4	2.0	6.4
4.0	2.3	6.3
3.8	2.6	6.4
3.7	2.6	6.3
3.9	2.5	6.4
4.3	2.0	6.3
4.7	1.6	6.3
5.4	0.9	6.3
5.7	0.6	6.3
5.8	0.3	6.1
6.1	0.3	6.4



19–23

Separating the Motion of the Center of Mass

The angular momentum of the body shown in Fig. 19 – 23, with respect to the point O, is the sum of the angular momenta of its mass elements. It is customary to express this in the following way: we take one representative mass element m_i with position and velocity vectors $\vec{\mathbf{r}}_i$ and $\vec{\mathbf{v}}_i$ respectively and write $\vec{\mathbf{L}} = \Sigma_i m_i \vec{\mathbf{r}}_i \times \vec{\mathbf{v}}_i$ where the letter Σ (sigma) with the subscript "i" stands for "sum of all quantities with the subscript 'i'."

At each instant of time we resolve the $\vec{\mathbf{r}}_i$ into the position $\vec{\mathbf{R}}$ of the center of mass and the position $\vec{\mathbf{r}}_i'$ of the mass element m_i relative to the center of mass, as shown in the figure:

$$\vec{\mathbf{r}}_i = \vec{\mathbf{R}} + \vec{\mathbf{r}}_i'$$

Similarly we resolve the velocity vector $\vec{\mathbf{v}}_i$ of each mass element into the velocity $\vec{\mathbf{V}}$ of the center of mass with respect to the fixed point O and the velocity $\vec{\mathbf{v}}_i'$ of the mass element relative to the center of mass:

$$\vec{\mathbf{v}}_i = \vec{\mathbf{V}} + \vec{\mathbf{v}}_i'$$

Inserting these expressions for $\vec{\mathbf{r}}_i$ and $\vec{\mathbf{v}}_i$ into the expression for $\vec{\mathbf{L}}$, we have

$$\vec{\mathbf{L}} = \Sigma_i m_i(\vec{\mathbf{R}} + \vec{\mathbf{r}}_i') \times (\vec{\mathbf{V}} + \vec{\mathbf{v}}_i')$$

Multiplying out the brackets, we obtain four sums:

$$\begin{aligned}\vec{\mathbf{L}} = {} & \Sigma_i m_i \vec{\mathbf{R}} \times \vec{\mathbf{V}} + \Sigma_i m_i \vec{\mathbf{r}}_i' \times \vec{\mathbf{V}} \\ & + \Sigma_i m_i \vec{\mathbf{R}} \times \vec{\mathbf{v}}_i' + \Sigma_i m_i \vec{\mathbf{r}}_i' \times \vec{\mathbf{v}}_i'\end{aligned}$$

Let us examine each sum separately. In the first sum the $\vec{\mathbf{R}} \times \vec{\mathbf{V}}$ relates to the center of mass and not to the individual mass elements. It can be taken outside the summation sign:

$$\Sigma_i m_i \vec{\mathbf{R}} \times \vec{\mathbf{V}} = (\Sigma_i m_i)\, \vec{\mathbf{R}} \times \vec{\mathbf{V}} = M\vec{\mathbf{R}} \times \vec{\mathbf{V}}$$

where M is the total mass of the body. The first term is just the angular momentum due to the center-of-mass motion. In the second term we can take $\vec{\mathbf{V}}$ outside the summation:

$$\Sigma_i m_i \vec{\mathbf{r}}_i' \times \vec{\mathbf{V}} = (\Sigma_i m_i \vec{\mathbf{r}}_i') \times \vec{\mathbf{V}}$$

Because each $\vec{\mathbf{r}}_i'$ is the position vector relative to the center of mass, the sum $\Sigma_i m_i \vec{\mathbf{r}}_i' = 0$. (This is the generalization of the case of two masses, for which

$$\frac{r'_1}{r'_2} = \frac{m_2}{m_1}$$

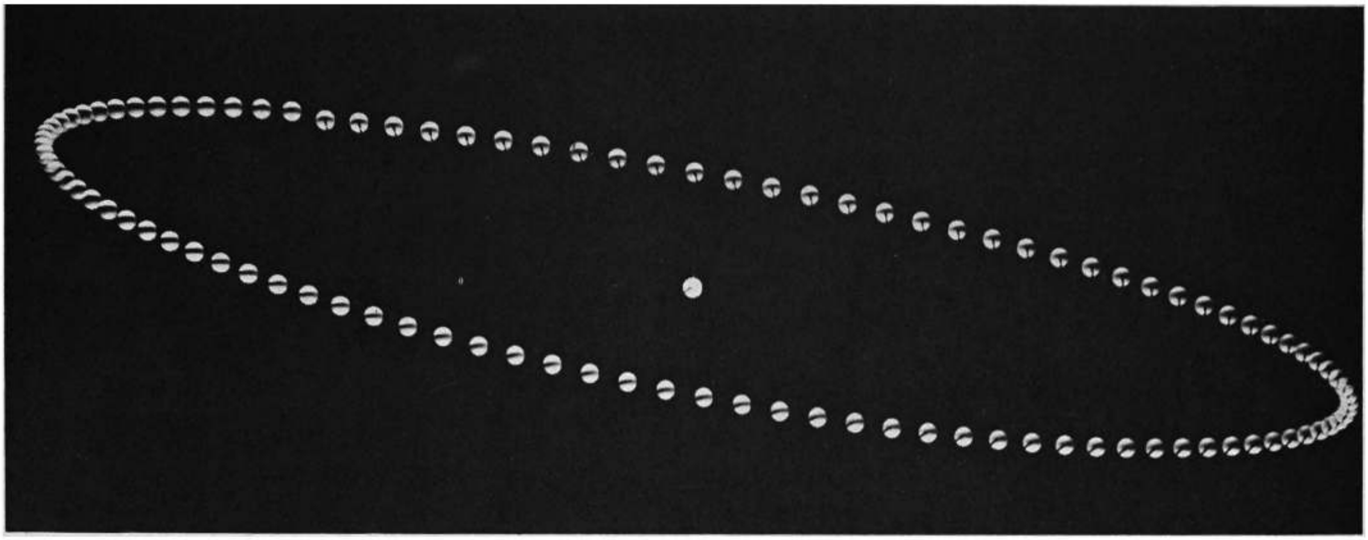
or, written vectorially

$$m_1 \vec{\mathbf{r}}_1' + m_2 \vec{\mathbf{r}}_2' = 0.)$$

In the third term we have the sum

$$\Sigma_i m_i \vec{\mathbf{R}} \times \vec{\mathbf{v}}_i' = \vec{\mathbf{R}} \times (\Sigma_i m_i \vec{\mathbf{v}}_i')$$

The sum $\Sigma_i m_i \vec{\mathbf{v}}_i'$ equals the rate of change with time of $\Sigma_i m_i \vec{\mathbf{r}}_i'$, which is zero at all times. Hence $\Sigma_i m_i \vec{\mathbf{v}}_i' = 0$. We are left with the last term $\Sigma_i m_i \vec{\mathbf{r}}_i' \times \vec{\mathbf{v}}_i'$. In this sum the position and velocity vectors are relative to the center of mass. This term, therefore, is the angular momentum due to the motion of the body relative to its center of mass, or the spin angular momentum.



19–24 For Question 4.

QUESTIONS

1.* Look at the two photographs in Fig. 19 – 1. Can you tell where the moving object was acted on (a) by a net attractive force? (b) by a net repulsive force? (c) What forces made up this net force?

2.* Does an object moving at constant speed in a straight line sweep out equal areas with respect to any arbitrary point in equal intervals of time?

3. If you had no information as how the photographs in Fig. 19 – 1 were taken, could you tell whether the moving body was subject to forces acting continuously or intermittently?

4. Examine the multiflash photograph in Fig. 19 – 24. From the photograph it is difficult to determine the velocity at the far ends of the orbit. How can you use the velocity at the point of closest approach to overcome this difficulty?

5.* At the instant shown in Fig. 19 – 4, the object is located at point B and the position vector $\vec{\mathbf{r}}$ is $\overrightarrow{OB}$. At a short time earlier, the position vector was $\overrightarrow{OA}$. What was the angular momentum relative to O when the object was located at point A?

6.* If two bodies have the same momentum $\vec{\mathbf{p}}$, must they have the same angular momentum L with respect to the same point?

7. An object is moving along a straight line with uniform velocity. What is its angular momentum relative to any point in space? Does this hold for an object moving in a circle with uniform speed?

8. A boy has a powered model airplane of mass 5.00 kg which is flying in a circle at 18.0 m/sec on a control line 15.0 meters long.

(a) What is the area swept out by the line per second?

(b) What is the angular momentum of the airplane?

(c) What is the force along the line required to keep the plane flying in the circle? (Neglect effect of tail; assume control line horizontal.)

9. Two vehicles A and B are moving along parallel roads in opposite directions at equal speeds. What is the angular momentum of A with respect to (a) a fixed point on its own road? (b) a fixed point on the other road? (c) the other car?

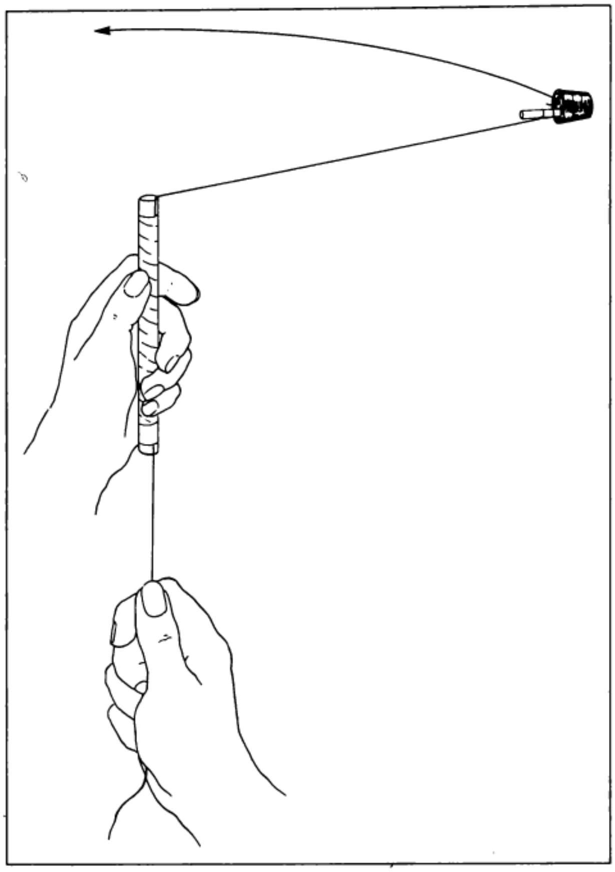
10. A boy whirls a rubber stopper on a string above his head as shown in Fig. 19 – 25. He slowly reduces the radius to one half by pulling the string with his other hand.

(a) How does this affect the angular momentum of the stopper?

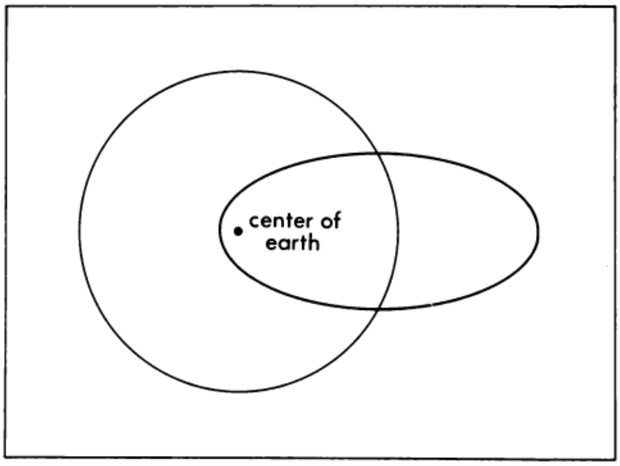
(b) The string breaks. Why?

11.* If two alpha particles move toward a nucleus with the same aiming error but from different directions, is it possible for them to have the same angular momentum relative to the nucleus?

12.* At what point in its trajectory near a nucleus is the angular momentum of an alpha particle equal to its momentum times its distance from the nucleus?



19–25 **For Question 10.**



19–26 **For Question 18.**

13.* (a) At which points in an elliptical earth-satellite orbit (Fig. 19 – 7) is $\vec{v}_r$ equal to zero?

(b) How do the perpendicular components of the $\vec{v}$ vectors compare at these points?

14.* Suppose that x_1 and x_2 are the two roots of the quadratic equation

$$ax^2 + bx + c = 0$$

(The roots are usually expressed in the form

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

where the positive sign yields x_1 while the negative sign yields x_2.) Show that the sum of the roots is

$$x_1 + x_2 = \frac{-b}{a}$$

and the product of the roots is $x_1 x_2 = \frac{c}{a}$.

15. Express both the potential and the kinetic energy of a planet at the end of the minor axis of its elliptical orbit in terms of its total energy only.

16. Set up at home a long pendulum, and try to get a circular orbit.

(a) Why is it easier to get elliptical orbits?

(b) Why do satellites describe elliptical orbits more frequently than circular ones?

17. For small displacements from its central position the restoring force on a pendulum is proportional to the horizontal displacement of the bob.

(a) How does the potential energy of the bob vary with the displacement?

(b) Make use of this to find the relationship between the major and minor axes of the elliptical orbit and the total energy and angular momentum of the bob relative to the center of force. (Hint: the center of force is at the center of the ellipse.)

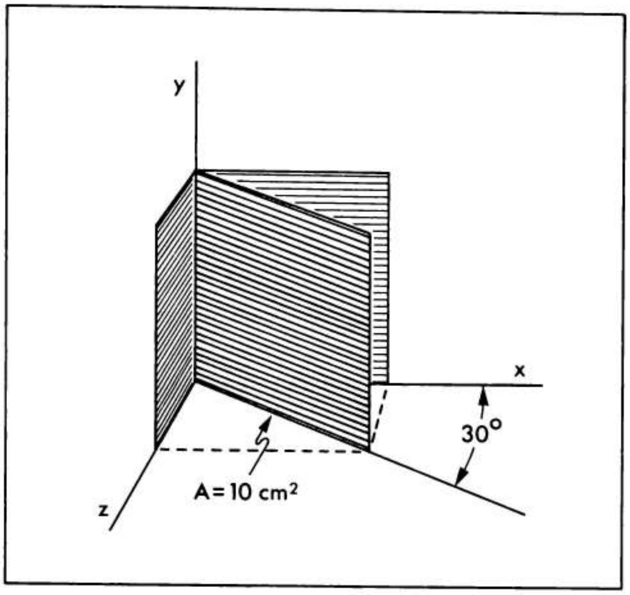
18. Figure 19 – 26 shows two possible satellite orbits, one circular and one elliptical. The semimajor axis of the ellipse is equal to the radius of the circle.

(a) How do the total energies of two identical satellites in these orbits compare?

(b) How do the angular momenta compare?

19.* In Fig. 19 – 27, a square of area 10 cm² is tilted so that it makes an angle of 30° with the x-y plane. Using the fact that area has the properties of a vector, calculate the area it projects onto (a) the x-y plane and (b) the y-z plane.

20.* What is the direction of the angular momentum relative to the center of force of (a) the alpha particle in Fig. 19 – 5? (b) the satellite in Fig. 19 – 7?



19–27 For Question 19.

21. A mass is moving in a horizontal circle on the end of a string.

    (a) Draw the angular momentum vector with respect to a point above the center for the two ends of a diameter of the circle.

    (b) Does the angular momentum vector stay constant?

22. Figure 19 – 28 shows a collision between a golf ball of mass 46 grams and a large wooden ball of mass 102 grams. Both are suspended from a point directly above the center of the white ring. Compare the angular momenta relative to the center of force of the two balls before and after the collision.

23. A golf ball and a putty ball of equal mass are suspended as in Fig. 19 – 9. They approach each other at equal and opposite velocities and collide.

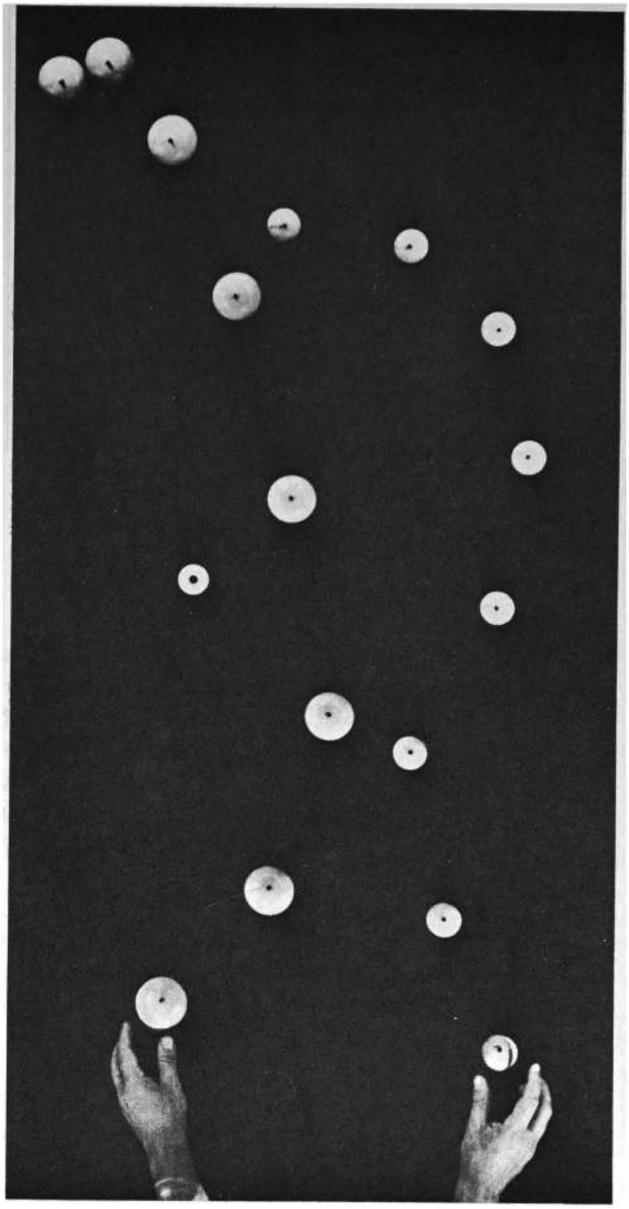
    (a) Describe their motion after they collide and stick together.

    (b) Is angular momentum conserved for this collision?

24.* What is the angular velocity in radians per second of a stroboscopic disc which is rotated 15 times in 10 seconds?

25. (a) How would you go about finding the total angular momentum of the puck shown in Fig. 14 – 11 just before the string attached to it is burned?

    (b) Is the total angular momentum conserved after the string is burned? How do you find out?



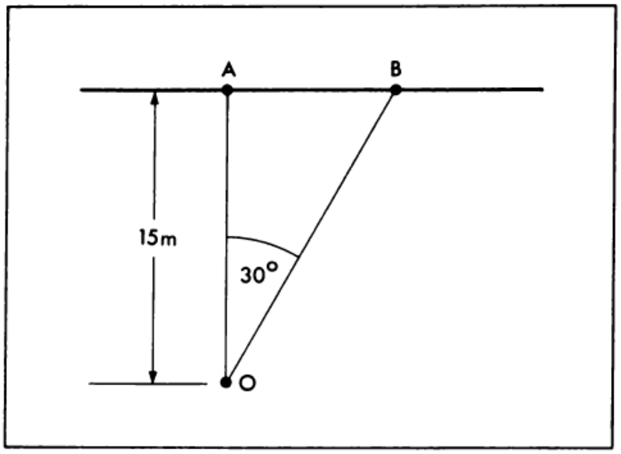
19–28 For Question 22.

26. It is occasionally suggested by science-fiction writers that the earth's rotation could be altered by lining up a huge number of automobiles and having them all accelerate toward the east at the same time. Suppose one did this experiment with 50×10^6 cars, each of mass 1.6×10^3 kg, all accelerating to 30 m/sec in 10 sec, near the equator.

    (a) How much angular momentum would be imparted to the earth?

    (b) Would any change in the earth's rotation be detectable?

27.* A thin rectangular sheet of metal measuring 6.0 cm × 8.0 cm is rotating about an axis through the center of the sheet perpendicular to the sheet at 2.0 revolutions per second. What is the speed of one of the corners of the sheet?



19–29 For Question 28.

28.* (a) A body moves in a circle of radius 15 m at a constant speed of 3.0 m/sec. What is its angular velocity with respect to the center of the circle?

(b) In Fig. 19 – 29, a body is moving on a straight line with a speed of 3.0 m/sec. What is its angular velocity with respect to the point *O* (i) when it is at point *A*? (ii) when it is at point *B*?

29.* Consider a dumbbell made of two equal bodies, each of mass *m*, connected by a rod of negligible mass and length *l*. What is the moment of inertia taken about an axis through the center of mass and perpendicular to the dumbbell?

30.* What radius must a ring have in order to have the same moment of inertia as a disc of equal mass and of radius *R*?

31.* Draw the angular momentum vectors for each of the three dumbbells in Fig. 19 – 18 one half a rotation later than the position shown in the figure.

32. Calculate the angular momentum of each of the three dumbbells in Fig. 19 – 18 about that point where the dumbbell rod intersects the axis of rotation. For each dumbbell, is your calculated angular momentum the same as the corresponding angular momentum relative to point *O* which is shown in the figure?

33. The moment of inertia of a thin uniform rod of length *l* and mass *m* taken about an axis perpendicular to the rod and through its center is $ml^2/12$. This can be derived by using integral calculus. Let us calculate an approximate value without calculus.

(a) As a first approximation, consider the rod to be made up of two mass elements, each of mass $\frac{1}{2}m$, located symmetrically with respect to the axis at a distance $\frac{1}{4}l$ from the axis. What result does this approximation yield?

(b) As a second approximation, consider the rod to be made up of four mass elements, each of mass $\frac{1}{4}m$, located symmetrically with respect to the axis, two at a distance $\frac{1}{8}l$ and two at $\frac{3}{8}l$. What result does this approximation yield?

(c) Calculate the moment of inertia of the rod by considering it to be made up of *n* mass elements, each of mass m/n, located symmetrically with respect to the axis with two at a distance $l/2n$, two at $3l/2n$, etc. Hint:

$$1^2 + 3^2 + 5^2 + \ldots + (n-1)^2 = \frac{n(n-1)(n+1)}{6}$$

34. The expression for the moment of inertia of a homogeneous sphere of radius *R* and mass *m* with respect to an axis through its center must have the form $I = kmR^2$ where *k* is a pure number. Why? Use your knowledge of the expression for the moment of inertia of a disc to estimate *k*.

35. (a) A figure skater can turn at 240 rpm with her arms pulled in. If her mass is 50 kg, what is her angular momentum? What assumptions are you making in your calculations?

(b) Now suppose the skater lifts and stretches her arms horizontally. Estimate her new frequency.

36. Figure 19 – 30 shows two lead cubes mounted to rotate horizontally around an axis passing through the hole in the center of the connecting bar: Estimate by how much (in percent) the moment of inertia of the two cubes differs from that of two equivalent "point masses" located at the centers of the cubes.

37. A constant torque of 30 newton-meters is exerted for 5 seconds on a flywheel initially at rest. The frequency of rotation at the end of this period is 120 rev/min.

(a) What is the moment of inertia of the flywheel?

(b) If the wheel is now brought to rest by friction forces in 5 minutes, calculate the torque provided by friction.

38. A man of mass 80 kg stands at the edge of a circular platform of radius 2.5 m which can rotate about a vertical frictionless axle. The moment of inertia of the platform is 1000 kg-m^2. The man now starts walking around the rim with a speed of 2.0 m/sec relative to the earth. What will be the effect on the platform? How long will it be before the man reaches the point on the platform from which he started?

39. An experimental bus used in Switzerland for comparatively short runs uses the energy stored in a flywheel which can be brought up to speed at the terminal points by an electric motor. The flywheel is coupled to a drive shaft by a specially designed clutch and gearing. From the following data calculate the minimum flywheel angular velocity required before leaving the terminal points.

    Mass of bus = 5000 kg

    Total resistance forces at normal speed (40 km/hr) = 2500 newtons

    Mass of rim-type flywheel = 500 kg (neglect spokes)

    Radius of flywheel = 1 m

    Distance between terminals = 10 km

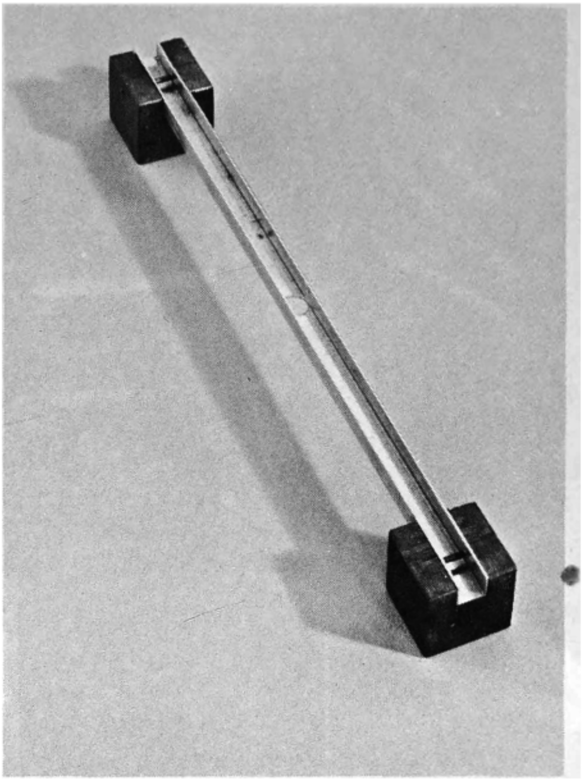
    An allowance of 50 percent above the uniform speed requirements is made to provide additional energy for starts, hill climbing, etc.

    The direction of the axis of the flywheel is important. How would different directions affect the riding and handling of the vehicle?

40. Spin a light circular disc on a nail. Blow down briefly on one side through a soda straw. What happens? Can you explain it by using the relation $\frac{\Delta \vec{L}}{\Delta t} = \vec{T}$?

41.* In Fig. 19 – 20, is the spin angular momentum in the same direction as the orbital angular momentum?

42.* In what units are the angular momenta in Table 2 expressed? Check one entry for both orbital angular momentum and spin.



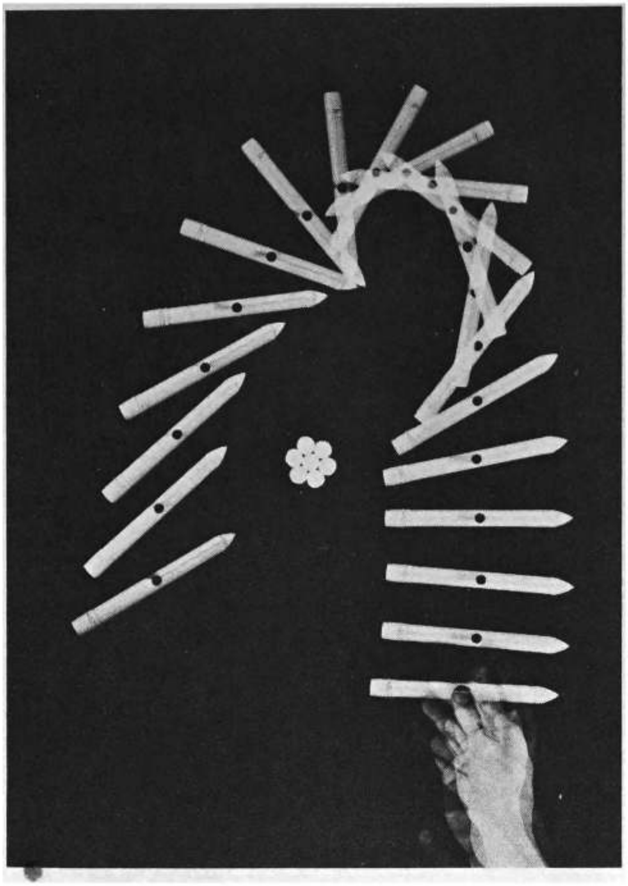
19–30 For Question 36.

43. (a) What is the angular momentum of the earth due to its rotation around its own axis? The mass of the earth is 6.0×10^{24} kg and its radius 6.4×10^6 m. What assumption have you made in your calculation?

    (b) What is the angular momentum of the moon about the earth? The mass of the moon is 7.3×10^{22} kg, its distance from the earth is 3.8×10^8 m, and its period of revolution is about 28 days.

    (c) What is the total angular momentum of the earth-moon system about the center of the earth?

    (d) An old astronomical theory proposed that the moon was originally part of the hot and liquid earth. To evaluate this idea, suppose that at a certain time the moon was a large "drop" on the surface of the earth, just about to be detached owing to the centrifugal force in the earth's frame of reference. What must have been the length of a day on earth just before the detachment? Is it likely the theory is correct?



19–31 **For Question 44.**

44. Figure 19 – 31 shows the motion of a magnet photographed under conditions similar to those in Fig. 19 – 20. Assuming that the fixed magnet has its north pole up, identify the poles of the moving magnet. Where in the figure is the net magnetic force repulsive? Where does the net magnetic force become attractive? Is the torque about the center of mass of the moving magnet zero at that point?

45. Take a pencil. Mark the center of mass on it. Place it on a smooth surface. Apply the thumb and first finger on either side of the center of mass, give a sharp twist, and pull your hand away. What happens to the pencil? Repeat this several times with both your thumb and first finger on the same side of the center of mass. Make sketch diagrams of what happens, showing motion before release, at moment of release, and following it. In which direction does the center of mass move? In which direction do the ends of the pencil move?

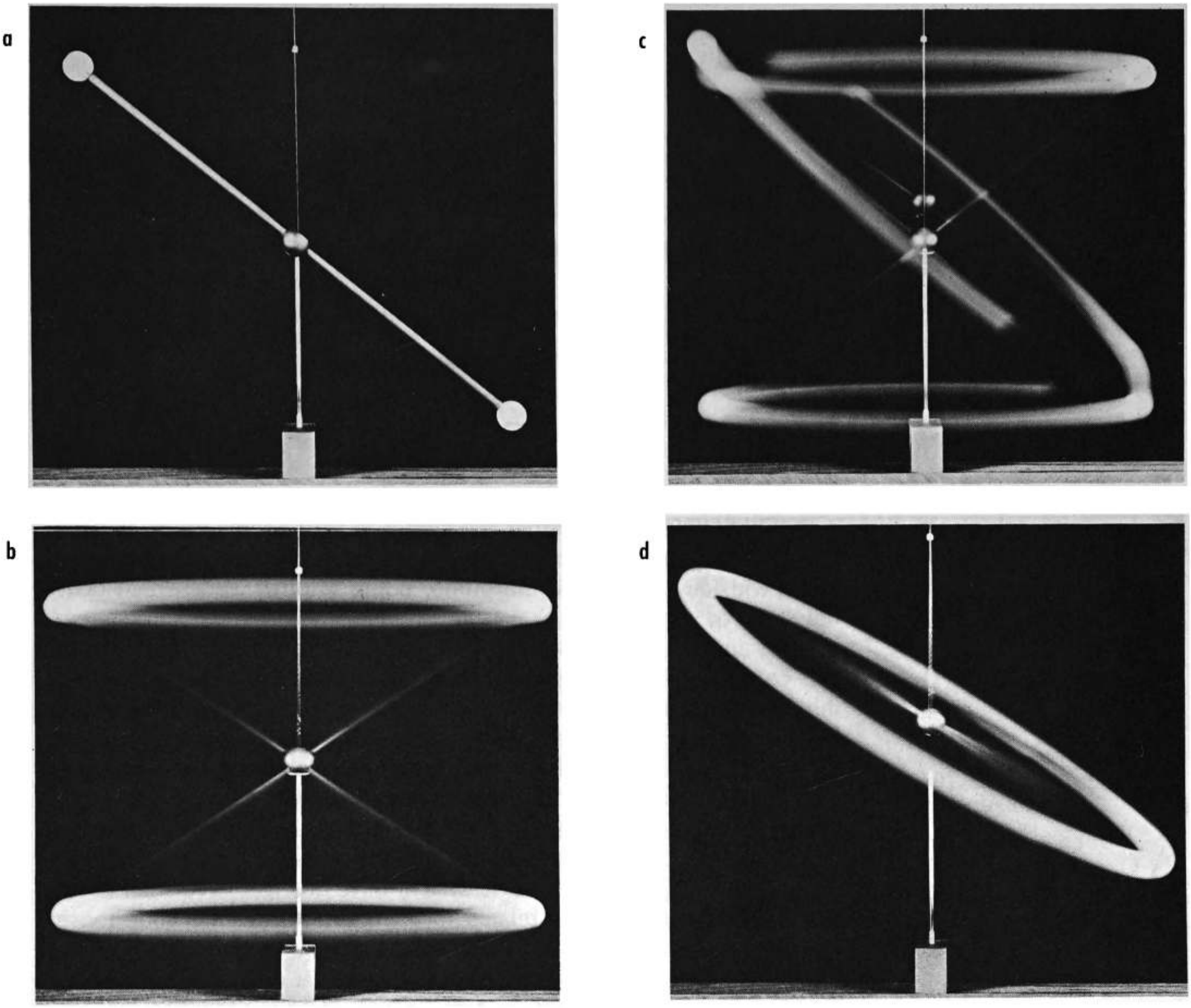
46. Two equal masses are attached to a string so that one mass is at one end and the other at the midpoint of the string. The free end of the string is now held and the masses rotated in a horizontal circle. The string is then released. Describe the motion of the two masses.

47. Figure 19 – 32 (a) is a photograph of a dumbbell like the one diagrammed in Fig. 19 – 18 (b), which is pivoted on a rod so that it rotates about a fixed vertical axis as seen in the time exposure in Fig. 19 – 32 (b). A string threaded through the dumbbell permits it to be lifted off the rod while rotating, as shown in Fig. 19 – 32 (c). Figure 19 – 32 (d) is a time exposure showing the motion after the dumbbell is separated from the rod and is supported only by the string. Explain the change in the motion of the dumbbell after it is lifted from the pivot.

48. (a) Two bodies *A* and *B* are acting on each other with equal and opposite forces along the line connecting them. Show that the total torque which they exert on each other with respect to any point in space vanishes. What does this tell you about the sum of their angular momenta?

    (b) Would your conclusion hold if the forces were equal and opposite but not along the line *AB*? Is it likely that such forces, depending only on the distances between the two bodies, exist in nature? Why?

49. The bob of a pendulum of mass *m* is swinging in a circular orbit about the vertical line passing through the point of attachment of the string to the ceiling. Eventually the pendulum comes to rest hanging along the vertical line. What happened to its angular momentum?



19–32 For Question 47.

20 Heat, Molecular Motion, and Conservation of Energy

A BOOK SLIDES along a table and comes to rest; energy seems to have disappeared, but the book and the table are slightly warmer. We fill a tire with air using a bicycle pump, doing a lot of work which does not seem to show up in any of the usual forms of mechanical energy, but the air and the tire and especially the pump do seem to heat up. We crush some ice in a bag by hitting it many times with a hammer; this time we do a lot of work and do not get a temperature increase, but we do find that some ice has melted as a result of our pounding. You could make a long list of such cases in which the work done on a system does not seem to balance with the changes in the mechanical energy of the system; the mechanical energy just seems to slip away. But in all such cases, if we look and measure carefully enough, we find that some other properties of the system have changed. We shall show that these changes can be related to energy in forms other than those we have yet considered, and that with these additional forms of energy included, we shall be able to retain the idea that energy is conserved. We shall begin with a simple system—a gas.

20–1 GAS PRESSURE, BOLTZMANN'S CONSTANT, AND MOLECULES

Two of the most striking properties of gases are (1) that they are easily compressed as compared with solids and liquids, and (2) that increasing the temperature of a confined gas increases the pressure it exerts on the walls of its container. All this can be summarized in a mathematical relationship derived from experiments with gases at low pressures and at temperatures well above the condensation temperature. For any sample of a gas, the pressure is

$$P = \frac{KT}{V}$$

where T is the Kelvin temperature (centigrade temperature + 273) and V is the volume.

TABLE 1

	DENSITY, D (gm/cm³)	MOLECULAR MASS, M (gm)	NUMBER OF MOLECULES IN A CUBIC CENTIMETER (D/M)
Carbon dioxide (CO_2)	1.977×10^{-3}	73.1×10^{-24}	2.68×10^{19}
Chlorine (Cl_2)	3.214×10^{-3}	117×10^{-24}	2.67×10^{19}
Helium (He)	0.1785×10^{-3}	6.65×10^{-24}	2.68×10^{19}
Hydrogen (H_2)	0.0899×10^{-3}	3.35×10^{-24}	2.68×10^{19}
Nitrogen (N_2)	1.251×10^{-3}	46.6×10^{-24}	2.69×10^{19}
Oxygen (O_2)	1.429×10^{-3}	53.2×10^{-24}	2.68×10^{19}

If we do experiments in which we determine K in the equation above for samples of different gases, all at the same temperature, pressure and volume, we find that K is proportional to the density of the gas. If we also measure the molecular masses of the molecules of different gases (a modern method for doing this is described in Chapter 28) we find that these masses also are proportional to the density of the gas. From these two facts we know that equal volumes of different gases at the same temperature and pressure contain equal numbers of molecules as can be seen from Table 1. From Table 1 we see that 1 cm³ of any gas at 0° C and atmospheric pressure contains 2.68×10^{19} molecules.

If we now write the expression $P = \frac{KT}{V}$, taking this into account we have

$$P = kT\left(\frac{N}{V}\right)$$

showing that the pressure of a gas is proportional to the Kelvin temperature and the number of molecules per unit volume. The new constant k is known as Boltzmann's constant and is a universal constant, the same for all gases.

The fact that equal volumes of different gases at the same pressure and temperature contain equal numbers of molecules was first proposed as a hypothesis by Avogadro in 1811 on the basis of chemical reactions between gases, long before it was possible to determine the number of molecules in a sample of a gas.

We can now determine the numerical value of Boltzmann's constant. We see from Table 1 that at atmospheric pressure and 0° C (273° K), there are 2.68×10^{19} molecules in a cubic centimeter of any gas, or $(2.68 \times 10^{19})(1.00 \times 10^{6}) = 2.68 \times 10^{25}$ molecules in a cubic meter. Therefore

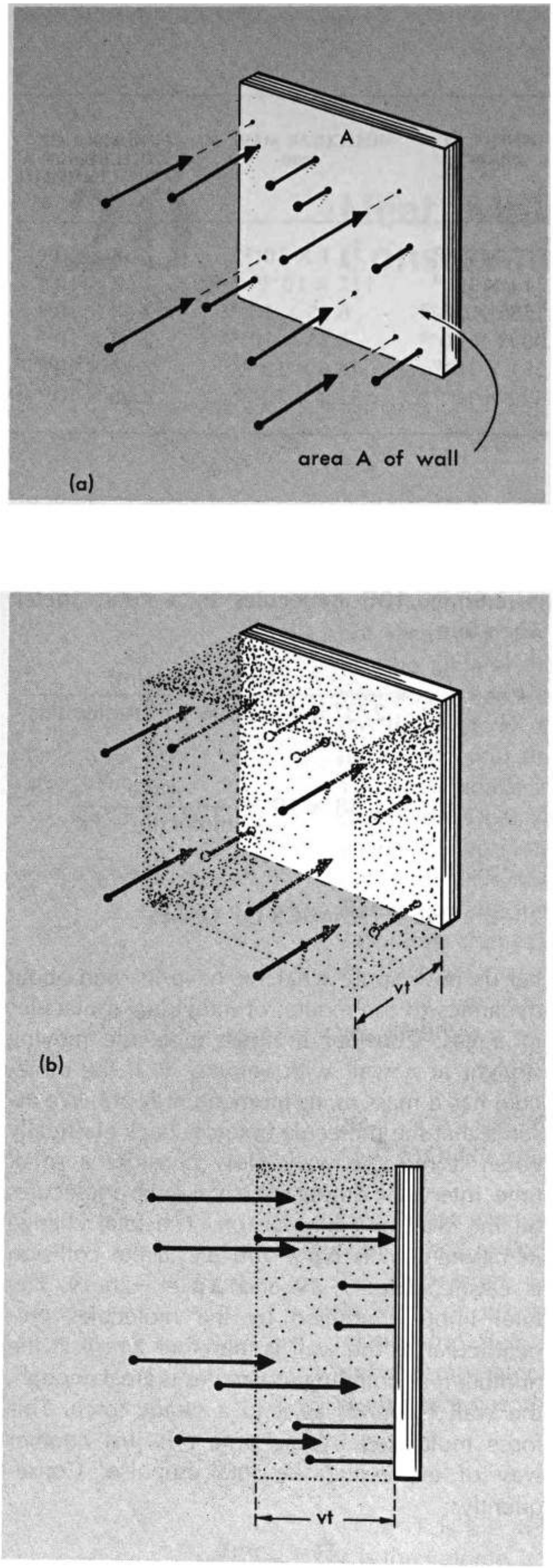
$$k = \frac{P}{T\left(\frac{N}{V}\right)} = \frac{1.01 \times 10^{5}\ \text{nt/m}^2}{273^\circ\ \text{K} \times 2.68 \times 10^{25}\ \text{molecules}}$$

$$= 1.38 \times 10^{-23}\ \frac{\text{joule}}{\text{molecule} - {}^\circ\text{K}}$$

20–2 THE DYNAMICS OF GASES

Let us now apply what we have learned about dynamics to the motion of individual molecules in a gas. Consider a single molecule moving straight at a wall with velocity $\vec{\mathbf{v}}$. If the molecule has a mass m, its momentum is $m\vec{\mathbf{v}}$. We assume that the molecule bounces back elastically when it hits the wall. Now consider a small time interval t during which n such molecules hit the wall and bounce off. The total change in momentum is $\Delta\vec{\mathbf{p}} = nm\,\Delta\vec{\mathbf{v}}$. If the collision is elastic, $\Delta\vec{\mathbf{v}} = -2\vec{\mathbf{v}}$ and $\Delta\vec{\mathbf{p}} = -2nm\vec{\mathbf{v}}$. The total impulse applied by the molecules perpendicular to the wall is therefore $2nm\vec{\mathbf{v}}$. If the number n of colliding molecules is great enough, the wall responds as if to a steady force. This force multiplied by the time t is just another way of expressing the total impulse. Consequently

$$\vec{\mathbf{F}}t = 2nm\vec{\mathbf{v}}$$



From this equation we see that the magnitude of the average force is

$$F = \frac{2nmv}{t}$$

We want to relate the force from this stream of molecules bombarding the wall to the force exerted on an area of wall by the molecular bombardment of a gas. For this purpose we wish to know how many molecules there must be in the little region of space near the wall to provide the bombardment we just considered (Fig. 20 – 1). We can find out by finding the volume near the wall from which the *n* molecules came.

Suppose we consider all the molecules that hit an area *A* of the wall during a short time interval *t*. Then all of them must have come from a little volume with the base *A* and altitude *vt*. No molecule that was farther away from the wall than *vt* can have reached the wall, because *vt* is the distance the molecules travel in time *t*, and if they started farther away they would fail to arrive during the time *t*. All the molecules that were approaching the wall from any distance less than *vt* must have reached the wall, because they would travel this shorter distance in less than time *t*. We can conclude, therefore, that at the beginning of the time interval *t*, the *n* molecules must have occupied the volume *vtA*. The number of molecules per unit volume that provide this bombardment is, therefore

$$\frac{\text{number}}{\text{volume}} = \frac{n}{vtA}$$

But this is only half the total number of molecules in a unit volume, since there are, on the average, as many molecules in the small volume *vtA* moving away from the wall as there are moving toward it.

20–1 (a) The molecules move toward the wall with velocity v. From what volume do the molecules come that reach the area A of the wall in time t? (b) The dots show the original positions of the molecules. The arrowheads show how far they get in time t – unless they hit the wall. Notice that all the molecules starting within the distance vt of the wall hit it. All those starting farther away do not get to the wall in time t. Therefore those molecules reaching the wall come from a volume of base A and height vt.

In a gas at uniform temperature, every identical volume of reasonable size contains almost the same number of molecules. If the gas is in a container, the number of molecules per unit volume is given by the total number of molecules N divided by the volume V of the container. In our model of a gas, therefore

$$\frac{2n}{vtA} = \frac{N}{V}$$

and

$$n = \tfrac{1}{2} vtA \left(\frac{N}{V}\right)$$

This equation relates the number of molecules n bombarding the wall area A in time t with the number of molecules per unit volume in a gas. It provides the connection for which we were looking. When we put this expression for n into the equation $F = 2nmv/t$ for the force on the area A of the wall, we obtain

$$F = \frac{2mv}{t}\left(\tfrac{1}{2} vtA \frac{N}{V}\right) = mv^2 A \left(\frac{N}{V}\right)$$

The average force per unit area is called the pressure P. According to our simplified gas model, then, the pressure is given by

$$P = \frac{F}{A} = mv^2 \left(\frac{N}{V}\right)^*$$

This is a very useful result, because the last equation now expresses gas pressure in terms of the number of molecules per unit volume and the kinetic energy per molecule.

In establishing this relation we have introduced all the essential factors in a molecular model of gas pressure. However, there are two final improvements we shall make in the model; we shall make the molecules move about in all directions instead of confining them to motion perpendicular to the walls, and we shall allow for the fact that the molecules have different speeds instead of assuming that all move at the same speed.

In a gas it is the random thermal motions of the molecules that bring them to the walls. The molecules strike the walls at all angles, not just perpendicularly. The net effect (as we show in the box on page 372) is to replace mv^2 by $\frac{1}{3}mv^2$. In a manner of speaking, only $\frac{1}{3}$ of the molecular motions are perpendicular to a single wall; the other $\frac{2}{3}$ of the motions are parallel to the wall. Consequently, the gas pressure is

$$P = \tfrac{1}{3}\,(mv^2)\,\frac{N}{V}$$

Now since $\dfrac{mv^2}{2}$ is the kinetic energy E_K,

$$\tfrac{1}{3}\,mv^2 = \tfrac{2}{3}\left(\frac{mv^2}{2}\right) = \tfrac{2}{3}E_K$$

We can therefore, write our expression for the pressure as

$$P = \tfrac{2}{3}E_K \left(\frac{N}{V}\right)$$

This form of the equation shows that the pressure is $\frac{2}{3}$ of the kinetic energy of a molecule times the number of molecules per unit volume. The equation emphasizes the role of the molecular kinetic energy in determining the pressure.

Finally, in a gas the molecular speeds are not all the same, and for different gases or mixtures of gases the molecules may have different masses. Here, however, we can tell from the last equation what to do. The equation shows that over a period of time, on the average, each molecule contributes to the pressure in direct proportion to its kinetic energy. The same equation will therefore give the pressure if we interpret E_K as the average kinetic energy per molecule. Multiplying the average by the number is the same as adding together the contributions of all the individual molecules.

It is important to remember that the kinetic energy involved here is the kinetic energy associated with motions of the centers of mass of the molecules. It is not the kinetic energy of the motion of the parts of the molecule vibrating or rotating about the center of mass. We can see that only the kinetic energy of center-of-mass motion is in question if we remember that the mv^2 arises from the momentum mv of each mole-

---

*You may wonder how the pressure, a scalar quantity, can be a force (vector) per unit area. Here, of course, we are concerned only with magnitudes, but the relation $F = PA$ holds also as a vector equation in the following sense. An area has the property of a vector whose magnitude equals the area and whose direction is along the perpendicular to the area. Then the force PA is also perpendicular and we have the vector equation

$$\vec{\mathbf{F}} = P\vec{\mathbf{A}}$$

cule times the speed v bringing the molecules to the wall. It is the speed of the molecules moving toward the wall that determines how many get to the wall per second and how big an impulse each gives to the wall. This is the center-of-mass motion, and the gas pressure therefore depends on the average kinetic energy in this motion.

Taking Account of the Random Directions of Molecular Motion

In Section 20 – 2 we show that a stream of molecules hitting a wall head on exerts a pressure on the wall

$$P = mv^2\frac{N}{V}$$

When the wall is bombarded by molecules coming at it in all directions, the pressure is

$$P = \tfrac{1}{3}mv^2\frac{N}{V}$$

We can account for the extra factor $\frac{1}{3}$ in the case of bombardment in all directions as follows: First, the components of molecular momentum parallel to the wall lead to no force on the wall. This is true because along every line parallel to the wall equal numbers of molecules move in opposite directions. And therefore equal numbers of molecules hit the wall with parallel components in one direction and with parallel components in the opposite direction. Now that we are rid of the parallel components, we look at the perpendicular components. In place of the momentum mv, we must now use the component $mv_\perp$; and in place of v for the speed which brings molecules to the wall, we must use $v_\perp$. Therefore in place of mv^2 we must now have $mv_\perp{}^2$. In other words, the square of one component of the velocity replaces the square of the velocity.

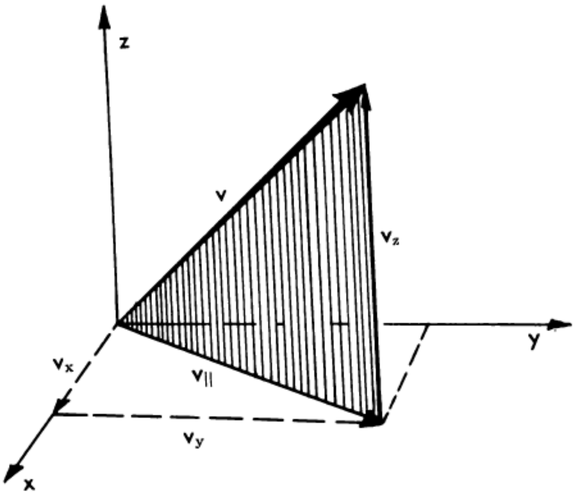
In the gas the velocities of the molecules must point equally often in all directions. We must therefore find the relation of the average of the square of the particular component $v_\perp$ to the square of the vector v when the vector points equally often in every direction. This can be done by expressing v^2 in terms of the squares of its components along three rectangular directions, say x, y, and z, where we have chosen z as the direction perpendicular to the wall. As Fig. 20 – 2 shows, $v^2 = v_x{}^2 + v_y{}^2 + v_z{}^2$. Now, when v points equally often in all directions, the average values of $v_x{}^2$, $v_y{}^2$, and $v_z{}^2$ must all be the same: that is, $\overline{v_x{}^2} = \overline{v_y{}^2} = \overline{v_z{}^2}$, where the bar means "average." Thus

$$v^2 = \overline{v_x{}^2} + \overline{v_y{}^2} + \overline{v_z{}^2} = 3\overline{v_z{}^2}$$

And since $v_\perp = v_z$,

$$v^2 = 3\overline{v_\perp{}^2} \quad\text{or}\quad \overline{v_\perp{}^2} = \tfrac{1}{3}v^2$$

This factor $\frac{1}{3}$ between $\overline{v_\perp{}^2}$ and v^2 accounts for the difference between the pressure of a stream of molecules hitting a wall head on and the pressure of molecules hitting a wall from all directions. Since the molecules in a gas move in all directions, the gas pressure is $P = \frac{1}{3}mv^2\frac{N}{V}$.



20–2 By the Pythagorean theorem, in the shaded right triangle $v^2 = v_\parallel{}^2 + v_z{}^2$. And $v_\parallel{}^2 = v_x{}^2 + v_y{}^2$ by using the same theorem on the right triangle with sides v_x, v_y, and $v_\parallel$. Therefore, $v^2 = v_x{}^2 + v_y{}^2 + v_z{}^2$.

20–3 TEMPERATURE AND MOLECULAR KINETIC THEORY; INTERNAL ENERGY

The molecular model, as we found in the last section, predicts that

$$P = \frac{2}{3} E_K \left(\frac{N}{V}\right)$$

where $E_K = mv^2/2$ is the average kinetic energy of the center-of-mass motion of a molecule in the gas. This is exactly the same as the relation discussed at the beginning of this chapter, $P = kT \; \frac{N}{V}$, but it goes further. It tells us that the proportionality factor kT is $\frac{2}{3}$ of the average kinetic energy of a gas molecule:

$$kT = \frac{2}{3} E_K$$

Because kT is the same for all gases at the same temperature, we learn that the average kinetic energy is the same, no matter what the mass of the molecule.

Now we can make our most important conclusion. Since

$$\frac{2}{3} E_K = kT$$

we find that the average kinetic energy of a molecule is $\frac{3}{2} kT$. Putting in the value of k, we get

$$E_K = \frac{3}{2}\left(1.38 \times 10^{-23} \; \frac{\text{joule}}{\text{molecule - K}^\circ}\right) T$$

$$= 2.07 \times 10^{-23} \; \frac{\text{joule}}{\text{molecule - K}^\circ} T$$

At room temperature, for example, a molecule has about 6×10^{-21} joule of kinetic energy on the average in its center-of-mass motion.

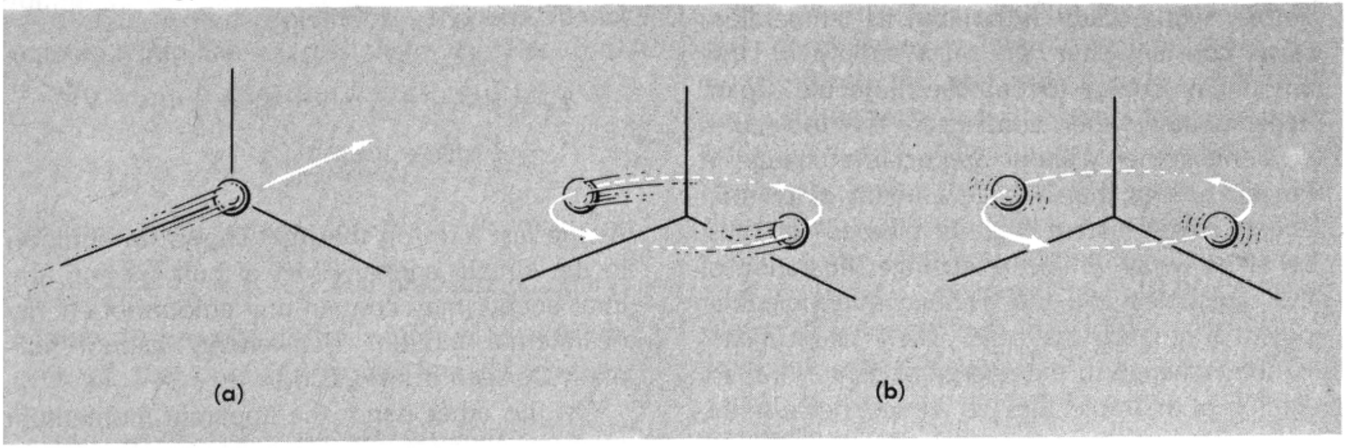
The temperature is a measure of the average kinetic energy of an individual molecule, the kinetic energy of its center-of-mass motion. One cubic centimeter of any gas at 0°C and a pressure of one atmosphere should have

$$(2.68 \times 10^{19})(2.07 \times 10^{-23}) = 5.54 \times 10^{-4}$$

joule of energy in this form for each Kelvin degree of temperature. In any gas, the molecules have this energy, and perhaps other forms of energy in addition.

We can put energy into a gas in many ways—beating it with an eggbeater, rubbing the container, letting heat flow into it are a few of the possibilities. (Here we do not mean throwing a bottle full of gas across the room—that gives energy to moving the whole bulk, while we are now interested in feeding the energy into the inside of the gas without moving the whole works.) When we feed energy into monatomic gases, it all appears to go into the energy of the motions of the centers of mass of the molecules. If we feed 5.54×10^{-4} joule into a cubic centimeter of helium or any other monatomic gas at 0°C and a pressure of one atmosphere the temperature rises by 1°K. We can put in the energy in many ways; all give the same result. The energy we put in is completely accounted for by the increased motion of the centers of mass of the molecules.

The molecules of the monatomic gases are especially simple. They are single atoms (He, Ar, Ne, etc.). More complicated molecules—like O_2, N_2, CH_4, which are made of two or more atoms—do not behave the same way. To raise their temperature one degree Kelvin may require more than 5.54×10^{-4} joule per cm³. Raising their temperature takes more energy because the atoms in a complicated molecule vibrate and rotate around the center of mass (Fig. 20 – 3). And when we feed in energy, only part of it ends up increasing the energy of center-



20–3 All the energy fed into a gas of monatomic molecules goes into center-of-mass motion of the molecules, as indicated in (a). Complex molecules, like those in (b), can also rotate and vibrate. Therefore, part of the incoming energy goes into rotation and vibration, and not all of it raises the temperature.

of-mass motion; the rest goes to increase the energy of rotation and vibration. Since the temperature measures only the part of the energy in the center-of-mass motion, more energy must be fed into the gas to produce a given rise in the temperature.

The energy of the motions of the centers of mass of the molecules and the energy of rotation and vibration of the atoms in a molecule are often called internal energy. Actually, not all internal energy is necessarily kinetic energy; it may include the mutual potential energy of the atoms as they change their separations inside the molecules.

When energy is fed into a gas, some of it may go into internal potential energy—inside the molecules. When the molecules are closer together, as in a liquid or solid, they continually exert appreciable forces on one another. Then, in addition to their kinetic energy, there is an average mutual potential energy of the interactions between the molecules; and when we feed energy in from the outside, some of it goes into changing the average potential energy of the interacting molecules. The internal energy then includes potential energy of molecular interaction.

The distinction between temperature and internal energy should now be clear. The temperature measures only the average energy of the center-of-mass motion. It gives the center-of-mass energy per molecule. The internal energy also includes the energy in all the other internal motions and even the potential energy which changes when the molecules move farther apart or closer together with changing temperature. The internal energy per molecule can therefore be different in two samples even though the temperature is the same. For example, while water is boiling, its temperature stays constant, but we must supply a large amount of energy to pull the molecules apart. When water vapor condenses, the molecules cluster together without appreciable change in kinetic energy, but a large amount of internal energy called "latent heat" is released and must be taken away while the average separation of the molecules and the average intermolecular potential energy decrease. The "latent heat" is not a change in the energy of center-of-mass motion of the molecules, as we see, because the temperature stays constant.

20–4 MECHANICAL ENERGY OF BULK MOTION AND INTERNAL ENERGY

When bodies isolated from the rest of the world interact with each other, their total momentum does not change. Momentum, as we saw in Chapter 16, is conserved. Mechanical energy, however, is conserved only if there are no dissipative forces such as friction. When there are frictional forces that depend on something besides the separations of the visible masses, the visible mechanical energy decreases, as in the example of a book sliding on a table. Is the law of conservation of energy really limited, unlike the law of conservation of momentum, or has the energy only changed form or location, so that its presence is no longer obvious?

Consider for a moment a closed box with two balls of equal mass m moving back and forth inside. Suppose that one ball moves at velocity $\vec{\mathbf{V}}$ when the other moves at velocity $-\vec{\mathbf{V}}$, and they pass each other at the midpoint. The system has no total momentum at any time, and from outside we see no motion. No kinetic energy is apparent unless we can look inside.

Now think of this whole system of box and balls moving north past us at velocity $\vec{\mathbf{v}}$ in the same direction as $\vec{\mathbf{V}}$, while the balls move back and forth inside with velocities $\vec{\mathbf{V}}$ and $-\vec{\mathbf{V}}$ relative to the box. The apparent kinetic energy $\frac{1}{2}Mv^2$ [where M is the mass of the box plus the mass of the two balls $(M_b + 2m)$] is only a part of the total kinetic energy. The total kinetic energy E_K is made up of the kinetic energy of the box M_b moving at speed v, plus the kinetic energy of the balls. If the box were transparent, you would see one ball m moving at speed $v+V$ and the other going at speed $v - V$. The total kinetic energy E_K is therefore

$$\begin{aligned} E_K &= \tfrac{1}{2}M_b v^2 + \tfrac{1}{2}m(v+V)^2 + \tfrac{1}{2}m(v-V)^2 \\ &= \tfrac{1}{2}Mv^2 + 2(\tfrac{1}{2}mV^2) \end{aligned}$$

As the last term on the right shows, in addition to the kinetic energy $\frac{1}{2}Mv^2$ of bulk motion, the total energy may contain any amount of energy of internal motion. That energy cannot normally be seen from outside.

On the other hand, the apparent momentum of the bulk motion *is* the total momentum. By adding the momentum of each mass, we get

$$M_b\vec{\mathbf{v}} + m(\vec{\mathbf{v}} + \vec{\mathbf{V}}) + m(\vec{\mathbf{v}} - \vec{\mathbf{V}})$$
$$= (M_b + 2m)\vec{\mathbf{v}} = M\vec{\mathbf{v}}$$

which is exactly the momentum of the total mass in its visible motion. These equations are very similar to the ones we developed in Sections 16 – 5 and 17 – 6 in our discussions of the center of mass. Indeed, in the present example, you recognize that the velocity of the box, $\vec{\mathbf{v}}$, is just the velocity of the center of mass of the system, and the velocities $+\vec{\mathbf{V}}$ and $-\vec{\mathbf{V}}$ of the balls are just their velocities relative to the center of mass.

This simple example illustrates a very common situation. Energy is often locked up in forms that do not appear as the kinetic energy of the motion of visible bulk masses, while momentum is less frequently hidden from our view. As a specific example of this, let us consider the collision of Fig. 16 – 14 between a golf ball of mass $m_1 = 45.7$ gm and a putty ball of mass $m_2 = 69.7$ gm. We saw, in our discussion of the center of mass of these balls, that the momentum of the balls was conserved throughout the collision. Before collision, the golf ball had a momentum $m_1\vec{\mathbf{v}}_1$; after collision, the two balls, stuck together, had a momentum $(m_1 + m_2)\vec{\mathbf{v}}' = (m_1 + m_2)\vec{\mathbf{v}}_c$. Thus

$$m_1\vec{\mathbf{v}}_1 = (m_1 + m_2)\vec{\mathbf{v}}_c$$

The golf ball brought into the collision a kinetic energy $E_K = \frac{1}{2}m_1v_1^2$; afterwards the kinetic energy of the visible bulk motion was only $E_K' = \frac{1}{2}(m_1 + m_2)v_c^2$. This is far less than the original kinetic energy, as we find by substituting $v_c = \dfrac{m_1}{m_1 + m_2}\, v_1$ from the momentum equation into the expression for E_K. The result is

$$E_K' = \frac{m_1}{m_1 + m_2}\, E_K$$

For the given masses of the balls, we find $E_K' = \dfrac{45.7}{115.4}\, E_K = 0.4E_K$. The final kinetic energy is only 40 percent of the initial value; 60 percent of the original energy has disappeared from view.

What has happened to this energy? We think that when the golf ball strikes the putty ball, it sets the molecules, both in the putty ball and in itself, into more rapid motion. Also, there is a permanent deforming of the shape of the putty ball, with accompanying changes in the potential energy of the molecules. A great deal of the kinetic energy of the golf ball has been transferred into the energy of random motions of the atoms and molecules and into changes in potential energy between these atoms and molecules. All of this is internal energy, hidden from our view. We lose track of this part of the energy and see only the fraction that goes into the bulk motion of the center of mass of the balls.

We would like to believe that no energy has really been lost; but if we can see only a very small part of the total energy, how can we know that all the energy is still around? We must find a way to measure the energy that has gone into internal motions or changes of atomic separations. In principle, perhaps, we could look at the position and motion of every microscopic or submicroscopic particle of matter; but in practical life we need some way of evaluating the energy stored in large chunks of matter without resorting to microscopic examination.

Here we are at the threshold of a great extension of the idea of conservation of energy. What we need is a measure of internal energy changes that depends on a few common measurements like temperature and volume. This will allow us to extend the conservation of energy to a whole new domain. With such a measure we can evaluate energy changes without seeing the actual position and motion of every microscopic particle of the object with which we deal.

We already have clues to the evaluation of the energy stored in internal motions and positions. When the golf ball collides with and sticks to the putty ball, the temperature of both balls increases. When a car is stopped suddenly, its apparent kinetic energy is lost; but the brakes warm up. When a meteor is slowed down in the earth's atmosphere, it gets so hot that it usually vaporizes completely. Where energy disappears from the visible motions and from the potential energy of the separation of visible bodies, we frequently notice a rise in temperature.

The temperature of a gas is a measure of the average mv^2 of its molecules. At least in the case of simple gases, the number of molecules affected times the change in temperature measures the energy which is fed into or out of the random internal motions. For other substances (for which we do not have such a simple model), the energy fed in may also go into changing internal potential energy, increasing the separations of atoms, for instance, as well as into the random motions. Nevertheless, as long as we keep the bulk volume of any substance the same, we might expect that the increase of internal energy would be reflected in a temperature change: a hotter body contains more internal motion than a colder one.

Now we can look back on the history of the idea of internal energy. Bacon, Galileo, Boyle, Hooke, Newton, and others conceived that the temperature of a body might be related to the "degree of motion" of the particles of which it was made. Boyle gave an illustration in the motion of a nail driven into a board. When the head prevents further motion of the nail, repeated blows of the hammer then serve to warm the nail: the motion of the hammer can no longer be imparted to the nail as a whole, Boyle said, and so is transferred to the "corpuscles" in the nail, making them move more quickly, and so producing a hot nail.

There were many contributors to the full experimental proof that internal energy must be included in the overall energy balance. Count Rumford observed the rise in temperature of cannon which he was boring. The mechanical energy supplied by horses was shown to be directly responsible for the temperature increase. Soon afterwards, the use of steam engines made it clear that at least some of the internal energy stored in fuels could be turned into mechanical energy. Quantitative demonstrations of the equivalence of various forms of energy were performed by the German physicist Julius Robert Mayer (1814 – 1878) and the British physicist James Prescott Joule (1818 – 1889).*

By the time of Mayer and Joule, there were available many processes of energy conversion: mechanical to internal energy, mechanical to electrical energy, internal to mechanical energy, electrical to internal energy, etc. Joule made measurements of *many* such conversions. In each experiment, he expressed his results in terms of *mechanical energy,* measured by raising a load against gravity, and *internal energy* (called "heat," in his day), measured by the increase in the temperature of a given amount of water. The constancy of the conversion factor between these two measurements, over the whole wide range of his experiments, firmly established "internal energy" as a real energy form. We shall now describe in detail one such experiment to show the equivalence between mechanical energy and internal energy.

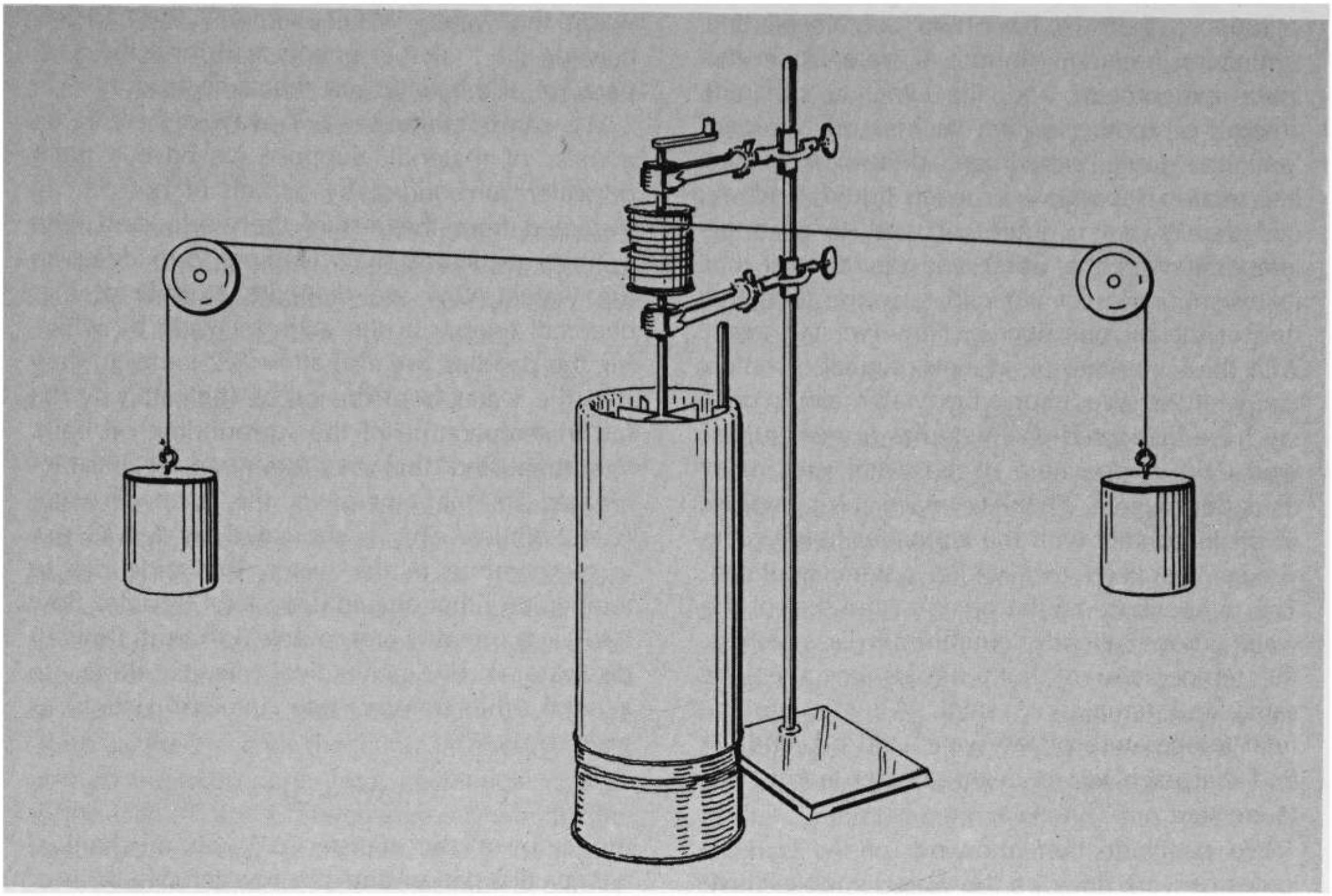
20–5 A QUANTITATIVE STUDY OF THE CONVERSION OF MECHANICAL TO INTERNAL ENERGY

James Prescott Joule showed that the same temperature change occurs in a fixed quantity of water—the same change in internal energy appears—whenever a given amount of mechanical energy is lost. Suppose that the water is placed in an insulated container; for example, in a large vacuum bottle. In the water there is a paddle wheel which can be turned by a falling mass m as shown in Fig. 20 – 4. When the bearings of the paddle wheel are well lubricated, the work done by the gravitational forces on the mass m measures the energy transferred from gravitational potential energy of m and used to churn up the water. When the mass is stopped, the paddle wheel stops, and the water quiets down. Finally it is at rest; but we find that the temperature of the water has increased. From beginning to end only two things have changed. The mass has fallen slowly through the distance h, so that mgh gravitational energy has disappeared, and the water has become warmer.

If mechanical energy and internal energy are equivalent energy forms, we shall always obtain the *same* temperature rise when a *given* amount of work is done in heating a given amount of water. In fact, a lot of different experiments have been done. They show that

---

*Mayer was a physician whose interest in the sources of animal energy led him to a study of energy in general. Joule, the owner of a successful brewery, was a pupil of Dalton. Joule made many distinguished contributions to science, expecially through his studies of electricity and energy conservation.



20–4 The apparatus for a Joule experiment, to observe the conversion of mechanical energy into internal energy. As the masses go down, the paddle wheel turns, churning up the water. When the masses have fallen the distance h, gravitational potential energy mgh has disappeared and the water has become warmer. (m = the total mass that goes down.) Note that the masses go down without attaining much speed, so their kinetic energy is small.

the temperature rise depends only on the energy transferred and not on the force or distance or the design of the paddle wheel. For example, if we double the mass of the falling body and allow it to fall only half the original vertical distance, the work done by the gravitational forces is the same as before. And we find that the temperature rise of the water is exactly the same, even though the process has been carried out considerably faster.

20–6 HEAT FLOW

The dissipation of mechanical energy through frictional forces is not the only method of increasing the temperature of an object. You can warm your hands either by rubbing them together or by dipping them into hot water. When they are in hot water, there is a transfer of energy from the more lively random molecular motions of the hot water to the more sluggish random molecular motions of your cold skin. When such a transfer of energy from a hotter to a colder body occurs, we say that *heat flows* from the first to the second body.

We can trace a given amount of energy from gross mechanical energy into internal energy of a given sample through either of two routes. The energy may be converted from mechanical to internal energy in the sample or it may be converted to internal energy elsewhere and reach the sample by heat flow. By experiment we can show that the method of transfer does not matter. The same amount of internal energy winds up in the sample either way.

Suppose that we have two containers, one containing a certain amount of water, as in the Joule experiment, and the other a different amount of another liquid, such as oil. We perform the Joule experiment described in the last section separately on each liquid; and we find that it takes x joules of work to raise the temperature of the water one degree and that it takes y joules of work to raise the temperature of the oil one degree. Now we start again with the water and oil at their original common temperature. We churn the water alone until we have dissipated $(x + y)$ joules of mechanical work. The temperature of the water rises more than one degree. Then we place the container of oil in contact with the container holding the water. We keep the whole system insulated. The temperature of the oil rises and that of the water drops. Finally equilibrium is reached: the temperature of both liquids becomes the same and remains constant. We measure the final temperature of the water and oil, and we find that each is one degree higher than it was at the start.

We conclude the following: of the $(x + y)$ joules of work done on the water, y joules have been transferred to the oil by *heat flow*. We make this conclusion because we know that exactly y joules of work are necessary to produce a temperature rise of one degree in the oil. The remaining x joules have produced the one-degree temperature rise of the water. Thus we have shown that the work necessary to produce a definite temperature rise in a given liquid is the same whether the energy is transferred to the liquid by heat flow or whether the mechanical energy is dissipated directly in the liquid.

20–7 QUANTITATIVE RELATION OF ENERGY DISSIPATION AND TEMPERATURE RISE

Let us be specific about the energy that must be dissipated to raise the temperature of a gram of water by 1°C. If we dissipate mechanical energy, we find that we must transfer 4.2 joules. Alternatively, we can put the energy in by heat flow from hot oil. We can calibrate the hot oil by a Joule experiment. Then we find that the oil must lose 4.2 joules by heat flow to warm the water. Whichever way they are received, 4.2 joules of energy will raise the temperature of the water just the same amount.

We can combine the two ways of heating up a mass of material. Suppose we have a gram of water surrounded by a bath of hot oil, all insulated from the rest of the world, and with a little insulating shaft leading to paddles in the water. Now we dissipate 1 joule of mechanical energy in the gram of water by whirling the paddles; we also allow 3.2 joules to flow into the water from the oil as indicated by the fall in temperature of the surrounding oil bath. We then find that the total transfer into increased internal energy of the gram of water is 4.2 joules. This is indicated by a 1°C rise in temperature in the water, the same rise in temperature that occurs if we let 4.2 joules flow into a gram of water purely by heat flow or dissipate it by mechanical means alone. In general, the transfer into internal energy is given by

$$\Delta W + \Delta Q$$

the sum of the transfer ΔW by mechanical energy dissipation and the transfer ΔQ by heat flow. This total transfer has the same effect no matter what fraction of it is ΔW and what fraction is ΔQ.

We can dissipate energy in water in innumerable ways—even dropping stones into it will do. The result is always the same. Whenever 4.2 joules are dissipated in a gram of water, its temperature rises 1°C. To raise the temperature of 2 grams takes just twice the energy. Similarly, whenever 12.4 joules are dissipated in 4 grams of helium, its temperature rises 1°C. The rise in temperature is different for different substances, as the examples show, but in any one substance we always get the same results over the same temperature range.

Precise measurements show that 4.185 joules are necessary to raise the temperature of 1 gram of water 1° from 14.5°C to 15.5°C. This amount of energy is called a calorie.* The calorie is just another unit of energy. We could measure the kinetic energy of the bulk motion of an object in calories just as well as in joules; for instance, a 3-kg mass moving 2 m/sec has $\frac{1}{2}(3)(2)^2 = 6$ joules or $\frac{6}{4.185} \approx 1.45$ calories of kinetic energy; however, the joule connects directly with our units of force and distance so that the joule is usually more convenient. We shall stick to joules. One unit is enough.

---

*Units like the calorie came into use very early in the study of heat flow. They are used because it is so easy to estimate the amount of energy being transferred through heat flow by transferring the energy to a known mass of water and measuring the temperature rise. Note that there is a larger unit, the "kilocalorie" or Calorie with a capital C, which is 1000 times as big as the small-c calorie. The Calorie is the unit commonly used today in specifying energy yields from food.

20–8 CONSERVATION OF ENERGY

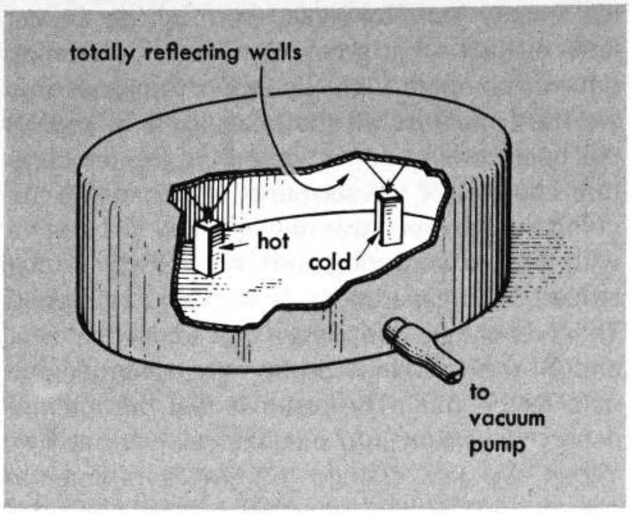
When gross mechanical energy is changed into internal energy, the same number of joules of mechanical energy gives the same number of joules of internal energy. If we put the energy into a standard amount of water, for instance, we always get the same temperature rise. But Joule experiments and heat-flow experiments fail to answer all the questions we can ask about internal energy and mechanical energy. Can we turn internal energy into mechanical energy at the same rate of exchange at which the mechanical energy turns into internal energy? Can we get 4.2 joules of gross mechanical energy from the internal energy that comes out of a gram of water when its temperature falls 1°C? Our earlier experiments suggest that we should get 4.2 joules, but we need direct evidence. Many things—even door latches—act one way when going in one direction and differently when reversed.

The simple way to check up on the change *from* internal energy *to* mechanical energy would be to pour a known amount of heat ΔQ into a heat engine and measure the work ΔW the engine does. We must be sure, however, to leave the heat engine at the end with the same energy it had at the beginning; otherwise we must know how much extra energy the engine has retained or how much it has given us from its original stock.

The usual way to make sure that the engine is not contributing energy is to get the engine back at the end into exactly the same condition it was in at the beginning. We then find, however, that the engine does not simply take in the heat ΔQ and produce work. It takes in the heat ΔQ, produces some work, *and* pours out some heat ΔQ. If we wish to see whether internal energy can be turned into gross mechanical energy joule for joule, we must use the net flow of heat—that going in minus that coming out—to compare with the work. Consequently, we must measure all the heat taken in and all the heat coming out by observing the temperature changes of the surroundings. Then we can compare the *net* flow of heat into the engine with the *net* work measured by forces times distances. Such experiments have been carried out. The Frenchman Hirn even traced all the heat and all the work in and out of the steam engine of a textile mill. The result is that the internal energy does turn into mechanical work at par. When we can change 4.2 joules of internal energy into mechanical energy, we get 4.2 joules of mechanical energy.

Although the equivalence of internal energy and mechanical energy is complete—the same when the transfer is in either direction—there is an essential difference between transfers in the two directions. We can easily change *all* of a supply of gross mechanical energy into internal energy; but we can change only a fraction of the available internal energy into gross mechanical energy if we do not allow any other changes to occur in the system. We pour heat into an engine at a high temperature; and normally some of it must be poured out again at a lower temperature; and the difference in internal energy turns into mechanical energy at full value. The fact that whatever energy is changed in form is converted at full value is a version of the first law of thermodynamics. The question of how large a fraction can be converted is the subject of the second law of thermodynamics, which will be developed in Chapter 22.

We now believe that internal energy and mechanical energy are just different forms of energy, and no energy is lost or gained in turning one into the other. Energy, like momentum, is conserved; but to see that it does not change amount we have had to find ways to measure heat flow and to keep track of the amount that is locked up in internal energies as well as the energy of bulk motions and raised weights. No single experiment leads all by itself to our conviction that energy is conserved. But we assume conservation in thinking about all kinds of physical and chemical processes. Innumerable such processes go on daily: chemicals react, heat is released, engines run. And energy bookkeeping balances out time after time. If any substantial



20–5 A hot object and a cold object are placed in a container with reflecting walls. There is a good vacuum in the container, so that energy is not transferred by the air molecules. Even so, the warm body cools off and the cold body warms up as energy is transferred by radiation.

amount of energy were disappearing, we should have felt the loss long since. If there were mysterious sources from which energy entered, the gain should have become apparent. It is on this overall check of conservation as well as on innumerable detailed checks of all sorts of transformations that our belief in energy conservation really rests.

By adding heat flow to mechanical work, we have extended the conservation of energy beyond simple mechanical systems. The extension was made by including a new transfer mechanism. Heat flow was the first new mechanism to be added to work, but we have since discovered that other mechanisms have to be included. For example, the energy reaching us from the sun does not arrive here either through mechanical work or through the kind of heat flow that occurs between a hot and a cold body in contact. Radiation—visible and invisible light—carries the energy. If we want the conservation of energy to span our universe, we must take account of this energy of radiation in the same way that we had to include the momentum of radiation to explain the behavior of bodies which emit or absorb light.

Experiment and theory combine to assure us that energy is indeed conserved when radiation is included. But, whereas the transfer of thermal energy by ordinary heat flow can be viewed as a microscopic version of the transfer of energy through work, the transfer through radiation is different. We cannot be satisfied that we understand that mechanism until we have studied radiation. Here we shall not try to go through the details. Even without studying the mechanism, however, we can extend the conservation of energy to include radiation by performing experiments like those with which we extended the conservation of energy to include heat flow. We can use a measured amount of energy to run a light for a while and absorb the emitted radiation on a black surface. The surface gets hot, and we can measure the energy which goes into heating the black surface and the energy left as heat in the source of light. We find the sum is equal to the energy we fed in. Part of the energy has gone from the light source to the absorber. And it has gone without loss or gain.

The radiation need not be visible. For example, we can put a hot body and a cold body inside of a very good reflector of radiation (Fig. 20–5). We make such a good vacuum inside the reflector that ordinary heat flow cannot take place. Nevertheless, the hot body loses energy and the cold one gains energy. The energy flows from one to the other as radiation. We find the total energy is conserved.

Like the transfer of momentum, the transfer of energy by radiation takes time. The light from the sun transfers its energy and momentum to us here 8.3 minutes after it leaves the sun. As far as we know, nothing ever moves faster. Energy transfer by other mechanisms is slower, but usually the distances are so short that we may fail to detect the time lag.

For many ordinary mechanical systems—such as billiard balls colliding or the earth going around the sun or even the operation of a steam engine—the part played by radiation in energy transfer is so small that we can usually ignore it. But radiation must be considered not only when we look into the energy transfer from stars but also whenever we consider atoms emitting or absorbing light.

All physical happenings, from the evolution of a star to the life of a firefly, are in essence transformations of energy from one form to the other. Even on a cosmic scale, we can trace its flow. Physicists have now learned about the nuclear energy that supplies the radiant energy of the stars. Our physical knowledge is consistent with a universal conservation of energy. To the best of our knowledge, the total energy in a region changes only when an exchange of energy occurs between the region and its surroundings. Then, if our region is the whole universe, we can expect no change of the total energy—because, by definition, there is nothing outside the universe. This assumption, that the total energy of the universe is constant, is the basis of most cosmological theories. But that is a daring generalization from our limited experience. We live in a small corner of the universe, and we have tested the validity of our physical laws over a very limited period of time. It is conceivable that these laws, including the law of conservation of energy, are not rigorously correct. In our limited view, energy *is* conserved with great precision; but a tiny rate of destruction or creation may have escaped our notice—if, for example, the total energy of the universe had doubled in the course of several billion years, we probably should not be aware of this change. Indeed, some cosmological theories assume a continuous creation of energy.

Here is one of the live scientific questions of today. Cosmologists are working hard to find out whether energy *is* completely conserved over the whole universe. And just now, for the first time, there is a possibility of actual experiments to decide between the rival theories.

QUESTIONS

1.* If a molecule of mass m and velocity v hits a wall head on and rebounds with velocity $-v$, what is the impulse given to the wall?

2. A machine gun fires a stream of 10-gram bullets at the rate of 400 rounds per minute. The bullets, moving at velocity 300 m/sec, hit a wall of solid rock and stop dead. Calculate:

   (a) the force on the wall

   (b) the kinetic energy of the bullets arriving at the wall in one minute.

   (c) the kinetic energy of the bullets in 1-meter length of the stream as they approach the wall. Compare twice this answer with your answer to (a).

3.* What is the pressure P on a wall having area A, which is pushed by force F?

4. Why is the pressure of a gas proportional to the number of molecules per unit volume?

5.* How does the total kinetic energy of the center-of-mass motion of the molecules in a sample of gas depend on the number of molecules N, the volume V, and the pressure of the gas P?

6. A gas in a cylinder pushes a piston out, increasing its volume by ΔV. The gas exerts a pressure P on the face of the piston, which has an area A, The force exerted by the gas moves the piston a distance Δx, transferring energy $F\Delta x$ to some outside machinery. Show that the work $F\Delta x$ is equal to $P\Delta V$.

   (*Note:* This is a very useful expression for work whenever we deal with a gas or a liquid pushing a piston; work equals the pressure times the change in volume.)

7.* By how much does the pressure of a gas in a steel cylinder increase if the temperature changes from 25°C to 100°C?

8.* By how much does the average kinetic energy of the center-of-mass motion of the molecules of a gas change if the gas is warmed from 25°C to 100°C?

9.* 10^{20} atoms of helium at 60°C are mixed with 10^{20} atoms of argon at 30°C. What is the temperature of the mixture?

10.* 10^{20} molecules of helium at 100°C are mixed with 10^{20} molecules of gas X at 20°C, and the final temperature of the mixture is 50°C. Can a molecule of gas X contain only one atom?

11. In a certain gas $\frac{2}{5}$ of the energy of the molecules is tied up in motion of the atoms around each other, and $\frac{3}{5}$ in motion of the centers of mass.

    (a) On the average, what is the kinetic energy of the center-of-mass motion of one such molecule when the temperature is 300°K?

    (b) If the temperature is raised 1°C, what energy must be supplied to 6.02×10^{23} molecules of the gas?

12. 6.02×10^{23} molecules of an ideal monatomic gas (in practice, helium or argon) is placed in a cylinder at temperature 273°K. The gas is at atmospheric pressure, 1.02×10^5 newtons/m². At this pressure and temperature the gas occupies 2.24×10^{-2} m³.

    A piston in the cylinder is then pushed in to decrease the volume by 2.45×10^{-4} m³.

    (a) How much mechanical work must be done to push the piston in? (Neglect the change in pressure.)

    (b) What is the final temperature of the gas if the container is completely insulated? (12.4 joules of energy raise its temperature one degree.)

    (c) By what fraction of its original value does the pressure change?

13. (a) Estimate the speed of oxygen molecules at room temperature from the following data: 32 grams of oxygen at room temperature (20°C) at one atmosphere pressure (1.02×10^5 newtons/m²) occupy 2.4×10^{-2} m³.

    (b) The same volume of hydrogen at the same temperature and pressure has a mass of only 2 grams. Estimate the average speed of hydrogen molecules at room temperature.

    (c) From your answer to (a), estimate the average speed of nitrogen molecules at room temperature, within 10 percent.

    (d) What is the average speed of air molecules at room temperature, within 10 percent?

    (e) What is the speed of oxygen molecules at room temperature and two atmospheres pressure?

14.* What is the total momentum of the molecules in a gram of helium gas in a container at a temperature of 300°K?

15. A large bag of sand is hung from a tree by a long rope. A boy shoots a bullet into the sandbag, and the bullet stays in the bag.

    (a) Describe the energy changes.

    (b) Suppose a 10-gram bullet is moving at 300 m/sec when it hits the bag, and the bag has a total mass of 1990 grams. Calculate the amount of kinetic energy:

    (i) the bullet had originally.
    (ii) the bullet and bag have after collision.
    (iii) that disappears.

    (c) What fraction of the original kinetic energy of the bullet goes into heat?

16.* In Fig. 20 – 4 why is it important for the masses to fall slowly?

17. In one of his most famous experiments, Joule churned water with a paddle wheel driven by two loads, each of mass 14 kg, each falling vertically about 2 meters. He had about 7 kg of water to be heated. After each churning he hauled his loads up and let them fall again. What temperature rise would you expect him to find after twenty falls?

    (*Note:* this is a reversal of the logic of Joule's great experiment. You know that 4.2 joules will always raise the temperature of a gram of water 1°C; Joule was trying to find this out.)

18.* A piece of lead is repeatedly hit with a hammer. Does its internal energy increase? Does heat flow into the lead from the outside?

19.* A hot stone is placed in a pail of cold water. Is work done by the stone on the water?

20.* If the pail in the preceding problem is thermally insulated, what happens to the total internal energy of the stone and water while the stone cools down and the water warms up?

21.* How many joules are required to change the temperature of 50 gm of water by 5 degrees centigrade?

22. Suppose we take 18 grams of water at boiling point and turn it into vapor in a cylinder. The cylinder is closed by a frictionless piston, which is so light that the gas remains at atmospheric pressure at all times.

    (a) What volume will the water vapor occupy if it behaves as an ideal gas? (Water vapor is not an ideal gas, but the error due to the assumption will be less than 10 percent.)

    (b) What work must be done in pushing the piston out against atmospheric pressure as the whole of the water vaporizes?

    (c) Given that each gram of water takes 536 calories to tear its molecules apart into vapor, how much heat is needed to convert the water to vapor and push the piston out?

    (d) What fraction of the total heat needed is converted to work in pushing the piston out?

23. A mass of 4 grams of helium at 24°C is placed in contact with water at 26°C insulated from the rest of the world. The final temperature of both is measured to be 25°C. How much water was there?

24. A high-speed swimmer uses 120,000 joules of energy in a half-minute race. Three quarters of the energy is wasted through heat flow; the rest

is dissipated by his hands and legs, by mechanical work.

(a) In 30 seconds he swims 50 meters. Estimate the average force opposing his motion.

(b) Describe the changes of form of energy in the swimming.

(c) Where and in what form is the energy that has been released when he has finished the race?

25. Suppose you have two large boxes of negligible mass. Box *A* contains 6×10^{23} molecules of hot helium at 60°C, the other box, *B*, contains 6×10^{23} molecules of cool argon (also monatomic) at 10°C.

(a) The two boxes are placed side by side in contact with all their outer surfaces insulated. After some time, both gases are at the same temperature. What is the temperature, and why?

(b) Instead of being placed side by side, the two boxes are joined together to make one large box, without change of volume, so that the gases mix. What will be the final temperature of the mixture?

(c) Now suppose that box *B* contains 6×10^{23} molecules of cold nitrogen at 10°C instead of the argon. The hot helium and cold nitrogen are allowed to mix as in (b) above. Will the final temperature be the same as in (b) or higher or lower?

(d) Give a clear reason for your answer to (c).

26.* Why do the brake drums get hot when a car goes at constant speed down a long, steep mountain road?

27. When you rub your hands together vigorously, they warm up but finally reach a maximum temperature no matter how long you rub. How do you account for the warming up and the eventual stopping of the temperature rise?

28. An amateur photographer had an electric fan in his darkroom to cool him when he was working there in hot weather. He would start the fan in the *closed* room 2 hours before he started to work there. Did he act wisely?

29. A mountain climber can climb about 500 meters vertical rise per hour.

(a) How much energy does such a climber gain as gravitational potential energy in a five-hour climb up a mountain?

(b) The human body is an inefficient chemical-mechanical machine. At best its muscles deliver only 25 percent of the chemical energy used as useful mechanical energy. The other 75 percent or more is wasted through heat flow. Assuming this efficiency, how much chemical energy does the climber use in 5 hours?

(c) Assume that, aside from mountain climbing, he needs 2.2×10^6 small calories per 24 hours. How much total energy should he take in his daily diet if he makes that climb every morning?

(d) If the climber walks down the mountain every afternoon, he loses the potential energy he gained. Why does that not help him to reduce his diet?

30. A rocket is equipped with insulated containers full of very hot gas instead of combustible fuel. The hot gas rushing out of a nozzle drives the rocket forward.

(a) Where does the momentum that the rocket acquires come from?

(b) Where does the rocket's kinetic energy come from?

(c) If the gas is ejected into a box standing on the ground as the rocket starts out, and the temperature of the gas collected in the box is measured after the rocket has left, do you expect that temperature to be higher, lower, or the same as the original temperature of the store of gas in the rocket?

31. (a) A cylinder containing helium gas is closed with a movable piston which has negligible friction. A man pushes on the piston and drives it in quickly, compressing the helium. The helium warms up. Why does the helium warm up? Discuss the mechanism of the warming up in terms of molecular behavior.

(b) A cylinder with a movable piston contains compressed helium. The piston is released and the helium pushes it out and cools. Explain in terms of molecular behavior how the helium cools.

(c) A large box with a good vacuum in it contains a small bottle of compressed helium. A trigger is arranged to remove the stopper of the helium bottle. When the helium is let out of the bottle, no change of temperature is observed after the release is all over. Explain, from the point of view of molecular behavior, why the helium does not change temperature when it expands into the big box.

(d) Although compressed helium shows no final change of temperature, some other gases show a noticeable cooling after they have expanded into vacuum from high compression. What does that tell you about these other compressed gases?

21 Irreversible Processes

21–1 SOME EXAMPLES

SUPPOSE YOU saw a motion picture showing bits of glass coming together all by themselves to form a perfect window pane, and at the same time a bullet receding from the pane to the muzzle of a gun. You have never seen this happen in real life. You would doubtless conclude that what you saw was the backward screening of a motion picture of a bullet shattering a window pane.

Our daily experience is full of examples of processes that seem to proceed in one direction of time but not in the reverse. They are generally called irreversible processes. Here are a few other examples.

Two balls of the same mass make a head-on collision with a given relative speed. After the collision they recede from each other at a lower relative speed. The mechanical energy lost is absorbed by the balls in the form of internal energy, and the temperature of each increases. We have never observed the reverse process, in which the temperature falls during the collision and each ball emerges at a relative speed higher than the initial one.

Two metal blocks, one hot and the other cold, are placed in contact. Gradually their temperatures become equal. We have never observed the reverse process, in which two metal blocks with the same initial temperature spontaneously change their temperatures, with one block becoming colder and the other becoming hotter.

A container is divided into two equal compartments. One compartment contains a gas, and the other is evacuated. A window is opened in the partition between the two and the gas expands freely, very soon filling the entire container. The gas is never observed to reverse this motion spontaneously so that it becomes confined to one compartment again.

As you know, all the foregoing processes obey the law of conservation of energy. If we imagine the reverse processes to take place, they could do so without violating the conservation of energy. In fact, they need not violate any other known conservation law, such as the conservation of momentum or of angular momentum. Since these reverse processes are not prohibited by conservation laws, why do they not occur?

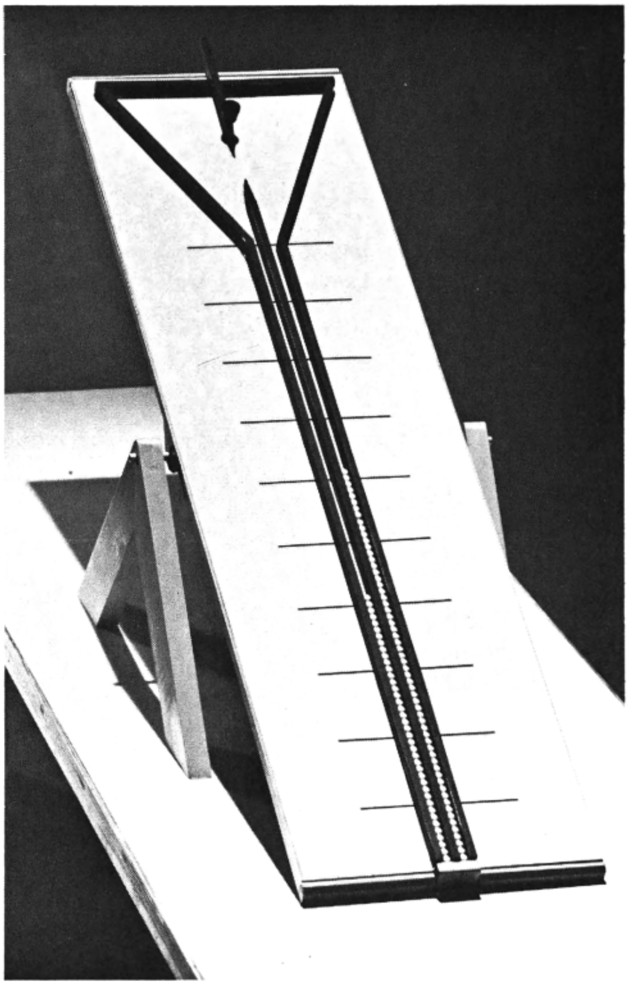
The last process described, that of a gas expanding into a vacuum, is one of the simplest examples of irreversible behavior. It will serve as a prototype for all irreversible processes. When the reason for irreversibility is understood in this example, only a slight extension is needed to enable us to understand it in the other examples as well.

The fact that irreversible processes do not violate known conservation laws probably means that the reverse processes may indeed occur; but for some reason the likelihood of their occurrence is so small as to be completely negligible. We shall show that such is indeed the case. The extreme rarity of the reverse process will be shown to be a consequence of the fact that macroscopic systems contain a very large number of atoms.

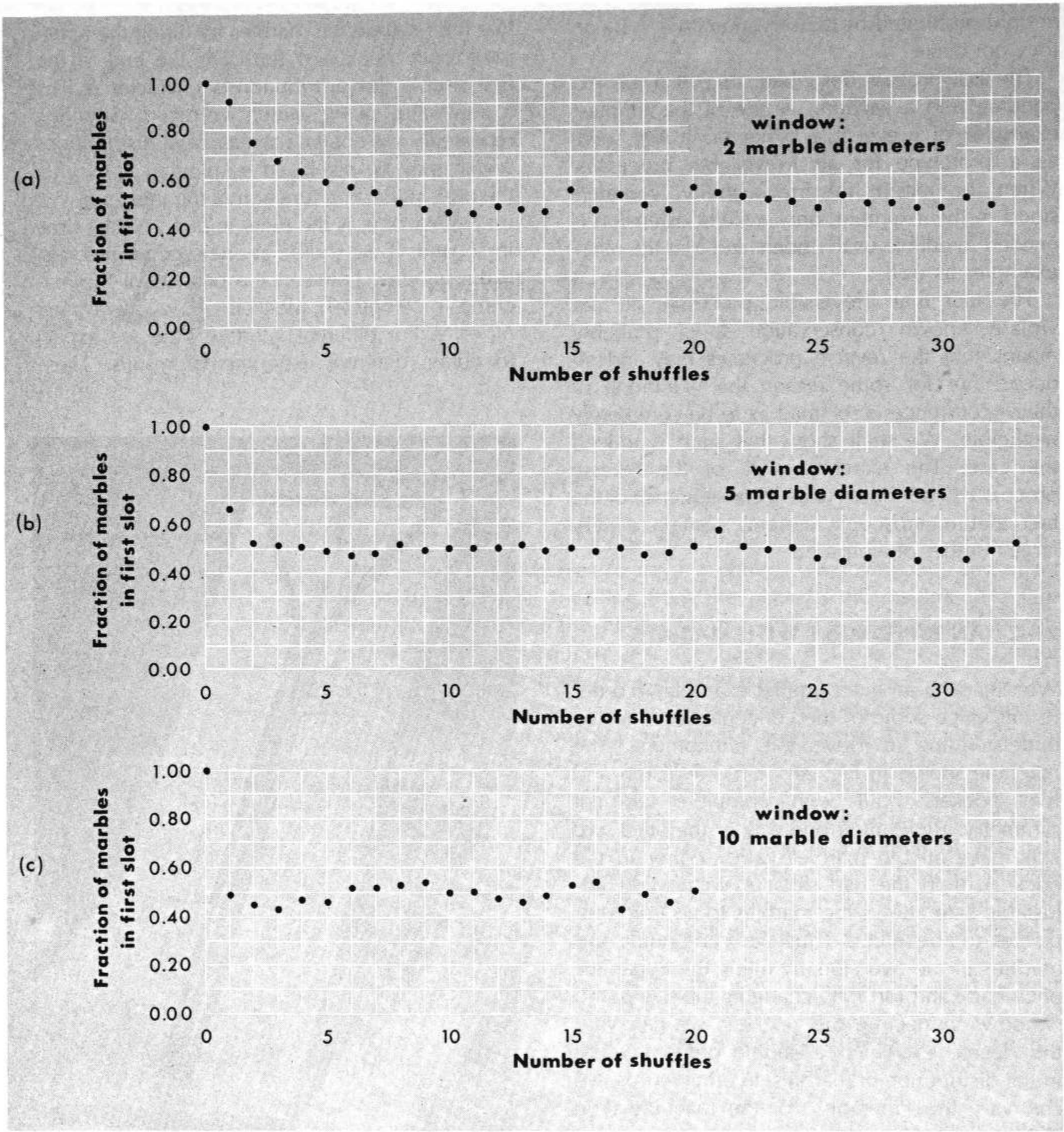
21–2 AN EXPERIMENT WITH MARBLES

We start with an even simpler example, in order to introduce some of the concepts useful for an understanding of irreversible phenomena. The machine shown in Fig. 21 – 1 has two slots and two pockets built with complete left-right symmetry. By tilting it one way or the other, we can make all the marbles gather either in the pockets or in the two slots. A window of adjustable size allows the marbles to go from one pocket to the other. Thus we can shuffle the marbles in the two slots by tilting the apparatus once back and forth. By changing the size of the window connecting the pockets, we can vary the degree to which a shuffle obliterates the initial distribution of marbles. In other words, we can vary the "memory" of the machine. The slots are made one marble-diameter wide so that the marbles will line up and can be easily counted. The *distribution* of the marbles is given by the number N_1 of marbles in the first slot, and the number N_2 of marbles in the second slot. The total number of marbles is $N_1+N_2=N$. Since N is constant, we need only to give N_1 to specify the distribution of marbles in the two slots.

We now try an experiment with 100 marbles. Initially they are all in the first slot, i.e., $N_1=N$. We then shuffle the marbles by tilting the apparatus once back and forth. At the end of the first shuffle the distribution is characterized by a new value of N_1, which we record. Shuffling repeatedly we obtain a sequence of distributions which may be displayed by plotting N_1/N after the nth shuffle as a function of the integer n. We may think of n as the time. In fact, n is the time in seconds if we shuffle at one-second intervals. If the window between the pockets of the machine is completely blocked off, we always get $N_1 = N$. For different window sizes we expect to obtain different experimental results. These



21–1 The marble machine, showing the two pockets at the top and the marbles in the two slots. Notice the "window" through which the marbles can pass from one slot to the other.



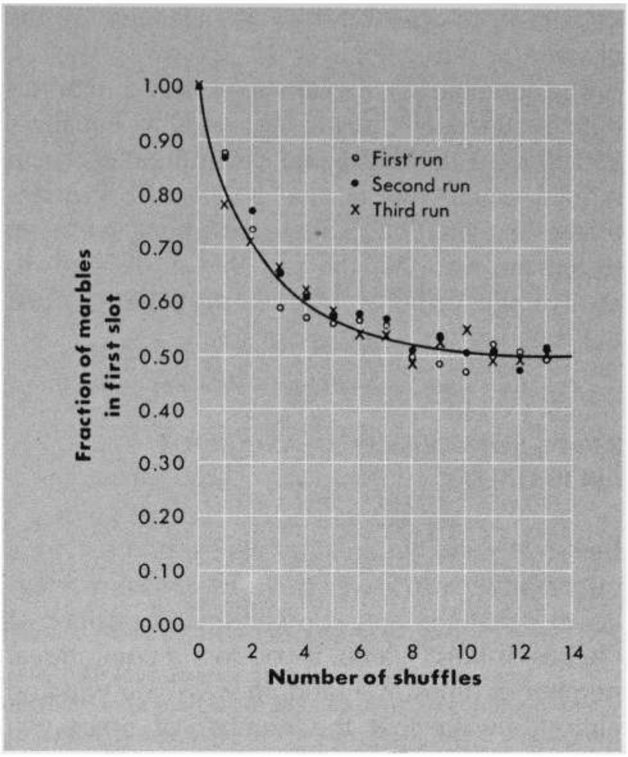
21–2 The results of three experiments in which only the window size was varied. All used 100 marbles, placed initially in the first slot.

are shown in (a), (b), and (c) of Fig. 21 – 2 for three sizes of the window. The experimental results show that as long as it is possible for marbles to go from one slot to the other, the initial distribution of marbles will tend toward an equal distribution. The time it takes for the equal distribution to be established depends on the size of the window. After the equal distribution has been established, it is not changed by further shuffling, apart from small irregular fluctuations. The transition from the initial unequal distribution to the final equal distribution apparently is irreversible.

Figure 21 – 3 shows in more detail how the distribution approaches equalization. The combined data from three different runs of the experiment all lie approximately on a single smooth curve. This shows that the distribution approaches equalization in a definite, reproducible way, apart from small fluctuations.

What would be the result if we started with, say, only three fourths of the marbles in slot 1? Without actually doing the experiment, we may guess that such an initial distribution will also tend to an equal distribution. For example, for the same window size as in Fig. 21 – 3, the experimental graph will be the same apart from the fluctuations, except that we must move the origin of the abscissa to the right.

We summarize our findings by the following empirical rule: Regardless of the initial distribution and the size of the window, the marbles tend to become equally distributed in the two slots, apart from small irregular fluctuations. The initial distribution and the size of the window merely determine the time it takes for equalization to be established.



21–3 Composite plot of three different runs, each with 100 marbles, with a window size of 2 marble diameters as in Fig. 21–2(a).

21–3 QUALITATIVE EXPLANATION OF THE MARBLE EXPERIMENT

A distribution may be changed by shuffling because it is possible for a marble to go from one slot to the other. The detailed process is, however, very complicated. The forces experienced by each marble come from gravity, friction, collisions with other marbles, collisions with the walls, etc. The slot into which a marble falls after shuffling is determined by all these many forces, of which we have no detailed knowledge. Even if we knew these forces precisely, we might still be unable to predict the exact outcome of a shuffle, because the complexity of the forces may defy a mathematical treatment of the problem. Thus, we cannot hope to explain every detail of the graphs in Fig. 21 – 2. It is reasonable to hope, however, that we can explain the trend to equalization without a detailed knowledge of the forces, because this trend is independent of such details as the size of the hole and the initial distribution. Such an explanation would satisfy our purpose, which is to understand this apparently irreversible trend.

A qualitative explanation is simple. First we start with all the marbles in slot 1. Since the two slots are geometrically alike, the only asymmetry is due to the larger number of marbles in the first slot. After the first shuffle, it is likely that a few marbles will be in slot 2, because each marble originally in slot 1 has a chance to go over during the shuffle. Of course, the number in slot 1 cannot increase, because there were no marbles in slot 2. In the next shuffle a few more marbles will probably move from slot 1 to slot 2, and perhaps a few from slot 2 will move to slot 1. Because the number in slot 1 is larger than the number in slot 2, we expect more to leave slot 1 than leave slot 2. The net result of the shuffle is then a decrease in the number in slot 1 and an increase in the number in slot 2. Thus, with each shuffle there is likely to be a net loss of marbles from slot 1 and a net gain to slot 2, until the distribution equalizes. Subsequent shuffling then tends to maintain the equalization, because the chance for slot 1 to lose marbles is balanced by the chance for it to gain.

The qualitative explanation shows that an initially unequal distribution tends to equalize, and that an initially equal distribution tends to remain equal. But to give a satisfactory quantitative explanation of the experiment, we need to explain not only the fact that N_1/N tends to an average value of 1/2, but also the fact that the irregular fluctuations are small.

21–4 SOME BASIC IDEAS ABOUT PROBABILITY

Before discussing the marble experiment in a quantitative way, we shall review our ideas about chance and probability. Consider a "heads-or-tails" game. If you toss a coin a great number of times in a random way, you will find almost always that the number of times you throw heads does not differ very much from the number of times you throw tails. If you toss the coin one thousand times, you may be practically sure that the frequency of the heads, that is, the number of times in which heads appear divided by the total number of the throws, will differ from $\frac{1}{2}$ by less than 0.1. If you consider cases in which the number of throws is greater, you can count on smaller differences. With an extremely large number of throws, the likely differences of the frequency from $\frac{1}{2}$ become negligibly small. We express this fact by saying that the *a priori* probability of heads and of tails are equal.

From the repeated tossing of a coin you learn something else. Consider, for instance, a set of four throws in which you know that the first three throws have given heads. What will be the outcome of the fourth throw? If you repeat a great number of sets of four throws, considering only these in which the first three throws give heads, you will find an approximately equal number of heads and tails in all the fourth throws, exactly as in the preceding case. The same thing happens if you consider the outcome of the n^{th} throw in sets of n throws in which the first $n - 1$ have any given prefixed result. We shall express this fact by saying that the probability of the outcome of any throw is independent of the results of any previous throw; different throws give rise to independent events.

With these notions, we can answer various questions about heads-and-tails games. For example, in what fraction of a large number of pairs of throws will we get two tails? We know that in very close to half the pairs of throws the first throw will yield tails. In this group of pairs, about half of the second throws will yield tails and the other half heads. Thus, in about $\frac{1}{4}$ of all the throws both members of the pair will show tails. We say that the probability of getting two tails in a pair of throws is $\frac{1}{4}$. Similarly, in two throws the probability of one tails followed by one heads is $\frac{1}{4}$. This holds also for the probability of one heads followed by one tails.

What is the probability of getting in two throws one tails and one heads, independently of their order? Out of N pairs of throws (N very large) about $N/4$ pairs will have tails followed by heads, and about $N/4$ pairs will have heads followed by tails. Therefore, you will have about $N/4 + N/4 = N/2$ pairs of throws with one heads and one tails. The probability of this event is therefore $\frac{1}{2}$.

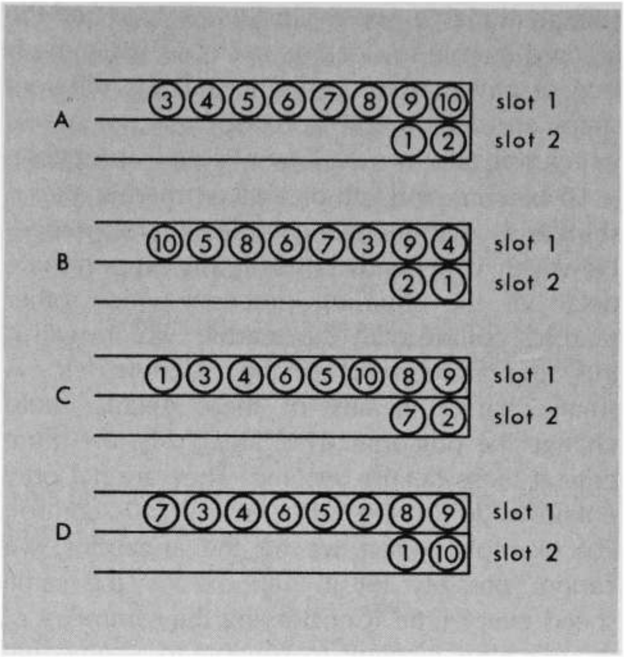
The method of reasoning which we have followed in deriving these results can be generalized as follows:

1. If you consider two possible events A and B which are mutually exclusive (that is, such that if A happens, B cannot happen, and vice versa) the probability of either A or B happening is the sum of the probability of A and the probability of B.
2. The sum of the probabilities of all the possible mutually exclusive events of a trial is 1, because one of the events must happen in every trial.
3. The probability that two independent events will both happen is the product of the probabilities of the two single events.

From these rules we can draw the following conclusions. If a trial has n (and only n) possible different outcomes, and if you know that all of the outcomes have equal *a priori* probabilities of happening, the probability of a given outcome must be equal to $1/n$. If you classify the outcomes of a trial in two different classes and if the number of events belonging to one class is m, the probability that an event belonging to that class will happen is m/n.

We have to bear in mind that the concept of "equal probability" of events has to be derived from experience. Once we have classified by experience all the possible different and mutually exclusive events in such a manner that they have equal *a priori* probability, we can apply the rules of probabilities for detailed calculations. The key problem, therefore, is to identify which events have equal *a priori* probability. It requires considerable care to avoid mistakes.

Consider, for instance, the problem of the outcome of a pair of throws in the heads-or-tails game. One may argue that there are just three possibilities: two heads, two tails, one heads and one tails. Assuming that they are equally likely, the probability of each would be $\frac{1}{3}$, which we have shown to be wrong. The mistake in this case is in having assumed as one event only, two different events (heads followed by tails, and tails followed by heads). We shall see an example of another kind of error in classification in the quantitative discussion of our marble experiment.



21–4 All four arrangements shown have the same distribution of marbles in the two slots. According to our definition, A and B are in the same state, but C and D are in states different from each other and from A's.

21–5 STATES AND DISTRIBUTIONS

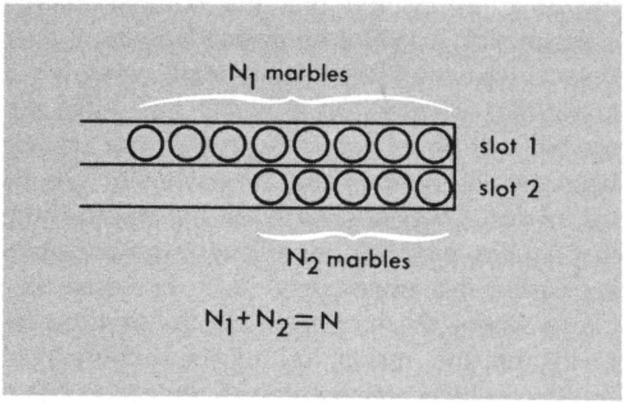
In order to achieve a quantitative description of the marble experiment, we have to find which events are equally probable. We may describe those outcomes in many ways. First, we can distinguish the N marbles by numbering them from 1 to N; the most detailed description then would be to specify not only which marbles are in each slot, but even the order of the marbles in each slot. Such a description is called an *arrangement.* Secondly, we may content ourselves in saying which marbles are in each slot regardless of their order; we shall talk then of a *state.* Third, we may give only the number of marbles in each slot regardless of their identity; this is called a *distribution* (Fig. 21 – 4). Notice that several arrangements may belong to one state, and many states may belong to the same distribution. An event will be the realization of an arrangement, a state, or a distribution. We know already from the reshuffling experiment that all the possible distributions are not equally probable. We have to look at either the states or the arrangements for equally probable events. To do so, consider first the behavior of one particular marble. Let us examine it experimentally.

For this purpose, let us repeat the reshuffling experiment. This time one of the marbles is identical to the others in all respects except color. (We may be sure that the color does not influence the result of the reshuffling.) If we follow the colored marble in a sufficiently great number of reshufflings, we shall find out in the long run the following two facts: The probability of finding the colored marble in a given lot is $\frac{1}{2}$, and this probability is independent of the manner in which the other (uncolored) marbles are distributed in each slot. To see that this is true, assume that the colored marble starts in slot 1. Will it be in slot 1 or slot 2 after n successive shuffles? With a small window we know that for $n = 1$ it is highly likely that the marble will be in slot 1. For $n = 2$, our answer becomes less certain, because whether it ends up in slot 2 depends partly on the result of the first shuffle, which is somewhat uncertain. As n increases, the uncertainty also increases, because where the marble is after n shuffles depends on the results of all the intermediate shuffles, which are themselves uncertain. For n

sufficiently large, we are just as likely to find the colored marble in slot 2 as in slot 1. The same is true of *any* marble, and since all the different states are determined by which marbles are in which slot, all states are *a priori* equally probable.

To be sure, the fate of a given marble after n shuffles is in principle completely determined. However, it depends critically on every minute detail of the shuffling process: which other marbles collide with our marble, which wall it hits, the air currents in the machine, etc. A small change in any of these details could change the outcome. The larger n is, the more critical these details become. They are not only unknown to us, but also beyond our control. For example, when we tilt the apparatus, we cannot possibly tilt it with exactly the same speed every time. Considering the symmetry of the machine, it seems reasonable to assume that after a sufficiently large number of shuffles a marble is equally likely to end up in slot 1 or slot 2. What constitutes a sufficiently large number depends on the size of the window and other details. It may be 10 or 100.

Notice that in this case the probabilities of consecutive shuffles are not independent as in the case of the heads-or-tails experiment. In our particular device the practically complete independence is realized only in the long run, because the results of immediately successive shuffles are not independent. But looking at the results of every, say, hundred shuffles, we can compare the marble experiment with the heads-or-tails game. We can in fact call the first slot "heads," the second slot "tails," and consider the dropping of different marbles as different throws of a single coin.



21–5 A distribution with N_1 marbles in slot 1.

Now, in the heads-or-tails game for N throws, all possible sequences of heads and tails are equally probable. For example, for five throws of a coin the sequence HHTTH corresponds in the marble experiment to marbles 1, 2, and 5 in the first slot and marbles 3 and 4 in the second. Such a description of the results of the marble experiment is a state. This analogy with the coin experiment confirms our earlier conclusion that all states in the marble experiment are *a priori* equally probable.

Notice that if we had assumed that all arrangements are *a priori* equally probable we would be in disagreement with experiment, since there are different numbers of possible arrangements for states belonging to different distributions.

21–6 THE MOST LIKELY DISTRIBUTION

What is the distribution of marbles after a large number of shuffles? A distribution, being specified only by the number N_1 of marbles in the first slot, in general does not correspond uniquely to a state. Many different states can have the same distribution. For example, the states C and D of Fig. 21 – 4 are different states, but they have the same distribution, i.e., $N_1 = 8$. According to the assumption of equal *a priori* probability, the most likely distribution should be the one that is shared by the largest number of states, because all states are equally likely to occur.

Let $W(N_1)$ be the number of states having the distribution N_1. To calculate this number, let us look at Fig. 21 – 5, which represents a distribution specified by N_1.

If a state having the distribution N_1 has been chosen, another state with the same distribution may be obtained by permuting the positions of the N marbles in Fig. 21 – 5. Not all permutations will do this, however. A permutation that rearranges only the marbles within each slot will leave the state unchanged. The number of permutations leading to a new state is the number of all possible permutations divided by the number of permutations that leaves the state unchanged.

What is the number of permutations of M objects? By a permutation we mean a rearrangement in the placement of the M objects with respect to some standard placement. There are

M possible choices for the position of the first object. After the choice is made for the first object, there remain $(M - 1)$ possible choices for the second object, and so forth. Finally, there is only one choice for the position of the last object. Hence the total number of all possible rearrangements is $M \times (M - 1) \times (M - 2) \times \cdots \times 2 \times 1 = M!$, which is called "$M$ factorial." We take $0! = 1$ by definition.

For the marbles in Fig. 21 – 5, the number of all possible permutations is $N!$. The number of permutations rearranging only the marbles in slot 1 but not in slot 2 is $N_1!$, and the corresponding quantity for slot 2 is $N_2!$. Since we may rearrange the marbles within each slot independently, the total number of permutations that rearrange the marbles only within each of the two slots is $N_1!N_2!$. Hence

$$W(N_1) = \frac{N!}{N_1!N_2!}$$

It is shown in Box 1 on p. 393 that $W(N_1)$ is largest when $N_1 = N/2$. This means that if we choose a state at random, it is most likely that $N_1 = N/2$, which is none other than the equal distribution. For this reason we refer to the equal distribution as the most probable distribution.

It is, however, by no means certain that by choosing a state at random we shall obtain the most probable distribution. We merely have a greater chance for obtaining it than any other distribution. On the other hand, the probability for obtaining the initial distribution $N_1 = N$ is small but not zero. The probability of obtaining any distribution N_1 is

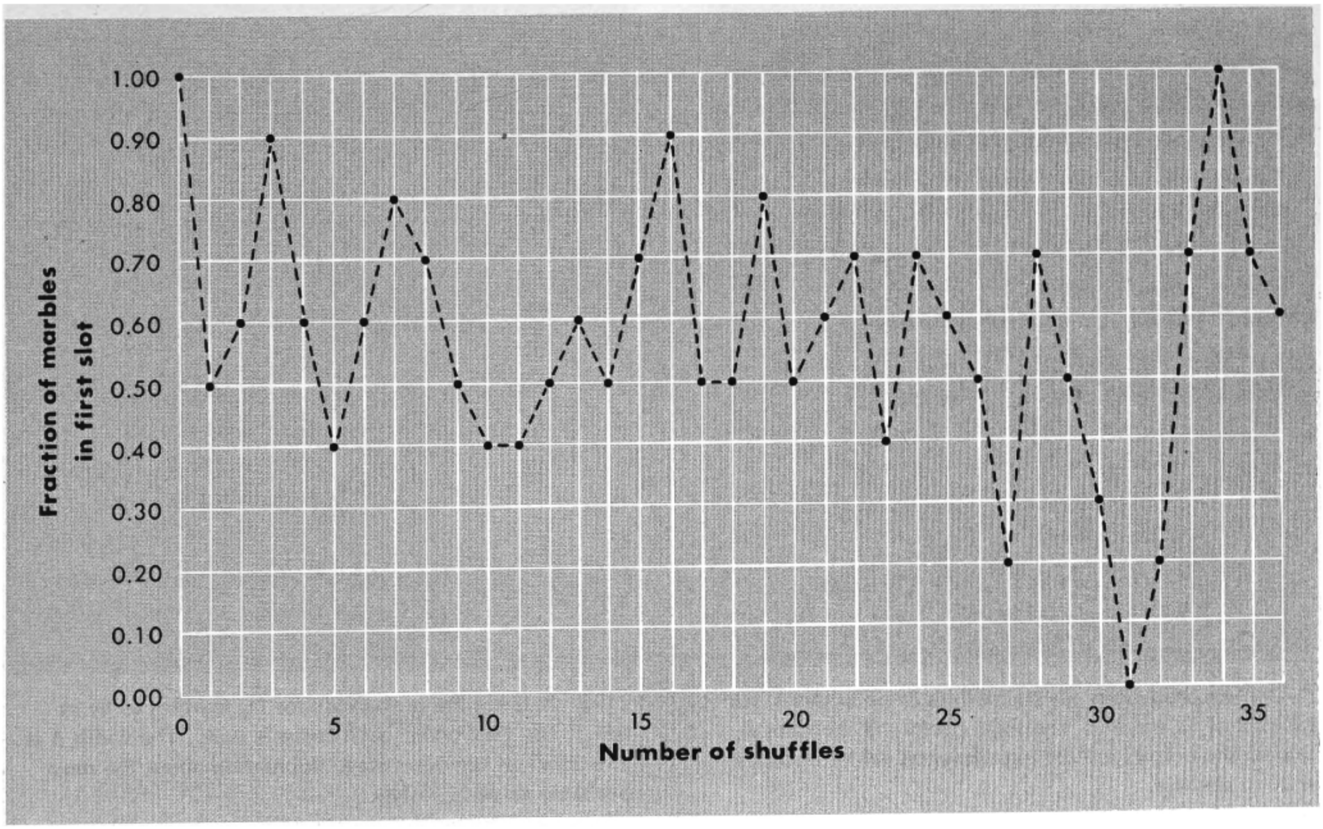
$$\mathcal{P}(N_1) = \frac{\text{Number of states with distribution } N_1}{\text{Total number of all possible states}}$$

$$= \frac{1}{2^N}\,\frac{N!}{N_1!N_2!}$$

The probability of observing the initial distribution $N_1 = N$ is

$$\mathcal{P}(N) = \frac{1}{2^N}$$

which decreases rapidly as N increases. For 10 marbles this probability is $\frac{1}{1024}$. That is to say, in an experiment with ten marbles, we can reasonably expect to get all 10 marbles in slot 1 once in 1024 shuffles. Assuming that we shuffled once every second for a long time, say a day, the initial distribution would recur at an average rate of once every 17 minutes. Figure 21 – 6 shows the result of an actual experiment with 10

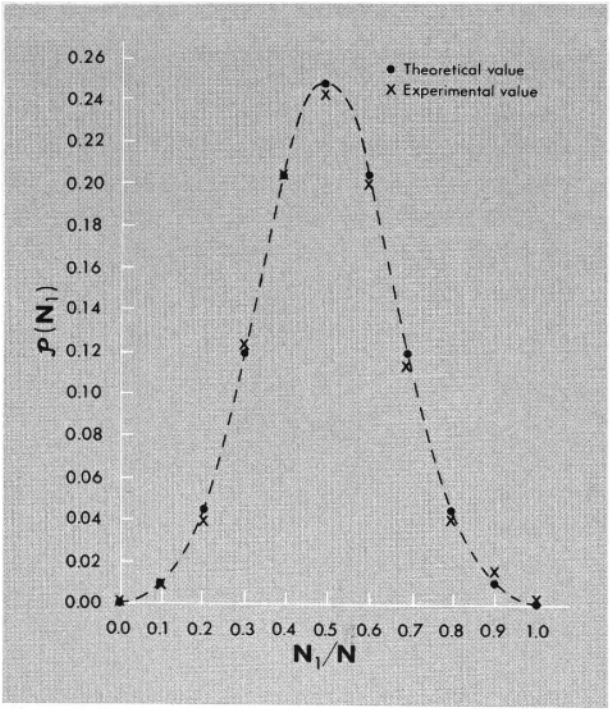


21–6 The result of an experiment with 10 marbles. The fraction of marbles in slot 1 is plotted against the number of shuffles.

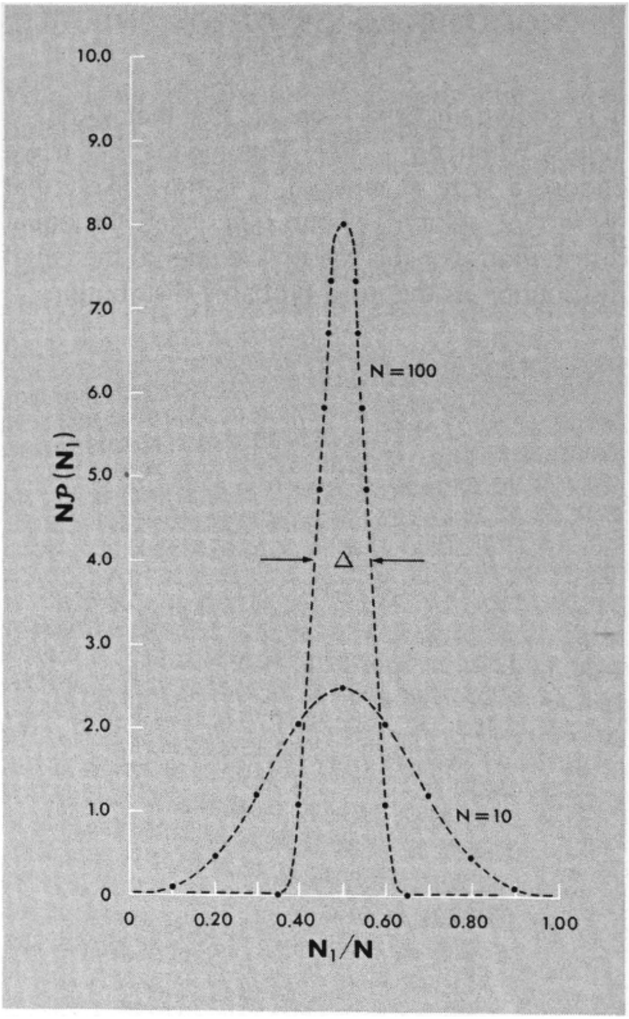
marbles. We see that although on the average $N_1/N \approx \frac{1}{2}$, the fluctuations are very pronounced. In fact, in this particular experiment the initial distribution recurs at the thirty-fourth shuffle. This does not contradict the previous prediction of 1024, because the latter is a prediction for the average behavior of a large number of shuffles.

For our experiment with 100 marbles, we would have to shuffle on the average 2^{100} times before we could reasonably expect to get all the marbles back in slot 1. At the rate of 1 shuffle per second, this would take 2^{100} sec $\approx 10^{23}$ years! This time is so long as to be physically meaningless, for the age of the universe is estimated at merely 10^{10} years. Before we would have time to see the recurrence of the initial distribution, the apparatus would have disintegrated into dust; the sun would have cooled; in fact, the laws of physics, for all we know, might have changed. That is why we can say that in the case $N = 100$ the chance of getting back the initial distribution is nil. The transition from the initial unequal distribution to the equal distribution is for all practical purposes irreversible. Of course, it is entirely possible that we get back the initial distribution $N_1 = N$ after a few shuffles. If this happens we would call it a freak accident, but it would not violate any natural law we know of. From a purely probabilistic point of view, however, any insurance company would gladly insure you against getting back the initial distribution at the odds of one dollar to the entire wealth of the world.

A convenient way to summarize all we have learned is to plot $\mathcal{P}(N_1)$ against N_1/N for different values of N_1. We do this for $N = 10$ in Fig. 21 – 7, where the theoretical points are shown as solid circles, and the experimental points obtained after 2000 shuffles are shown as crosses. We expect the agreement between them to improve if we increase the number of shuffles.



21–7 The probability for N_1 marbles to be in slot 1, for the case of 10 marbles. The solid circles are theoretical values. The crosses are the experimental values obtained in 2000 shuffles.



21–8 N times the probability for N_1 marbles to be in slot 1. The area under each curve is unity. The width Δ is a measure of the percentage fluctuation about the most probable value $N_1 = N/2$.

You can make a similar plot showing a comparison between theory and experiment for $N = 100$.

To exhibit the difference between the cases $N = 10$ and $N = 100$, thereby emphasizing the dependence of the fluctuations on N, it is best to plot $N\mathcal{P}(N_1)$ against N_1/N. The plots for $N = 10$ and $N = 100$ are shown superimposed on each other in Fig. 21 – 8. If we imagine the successive points connected by a line, the area under each curve is unity. We observe that the curves are both peaked at the most probable distribution $N_1 = N/2$, but the case $N = 100$ has a much sharper peak. The sharpness of the peak may be measured by the width of the peak at half maximum, denoted by the Greek letter Δ. It can be shown (Box 2, p. 394) that in general

$$\Delta \approx \frac{1.3}{\sqrt{N}}$$

Thus the width for $N = 100$ is about $\frac{1}{3}$ as great as the width for $N = 10$. For each curve most of the area under the curve lies within the width of the peak. This means that a distribution lying on the abscissa within a distance $\Delta/2$ of the most probable distribution has a large probability of occurring, while those falling outside of this range are improbable. In fact, the distributions close to the most probable distribution are quite likely to occur, because their probability is not much different from that of the most probable distribution. They show up in our experiment as the irregular fluctuations shown in Fig. 21 – 2. These fluctuations can be made smaller by increasing N, but they cannot be completely eliminated as long as N is finite.

1. *The Maximum of $W(N_1)$*

Consider the ratio

$$\frac{W(N_1+1)}{W(N_1)} = \frac{N!}{(N_1+1)!(N-N_1-1)!} \times \frac{N_1!(N-N_1)!}{N!}$$

$$= \frac{N-N_1}{N_1+1}$$

As long as this ratio is greater than 1, $W(N_1)$ is increasing and has not yet reached its maximum. But as soon as it becomes equal to or less than 1, $W(N_1)$ is either at its maximum or has already passed it and is decreasing again. This will happen when

$$N - N_1 \leq N_1 + 1$$

or

$$N_1 \geq \frac{N}{2} - \frac{1}{2}$$

$W(N_1)$ has a maximum for the smallest integer N_1 satisfying the above relation. If N is even, $N/2 - \frac{1}{2}$ is not an integer, and the smallest integer satisfying the relation is $N_1 = N/2$. If N is odd, $N/2 - \frac{1}{2}$ is an integer, and the smallest integer satisfying the relation is $N_1 = N/2 - \frac{1}{2}$. But in this case the next integer, $N_1 = N/2 + \frac{1}{2}$, actually gives the same value for $W(N_1)$. Thus if N is even, $W(N_1)$ has a maximum at $N_1 = N/2$, and if N is odd, $W(N_1)$ has two equal maxima at $N_1 = N/2 - \frac{1}{2}$ and at $N_1 = N/2 + \frac{1}{2}$. Note that if N is very large, the difference between $N/2$ and $N/2 \pm \frac{1}{2}$ is negligible, and we can say that $W(N_1)$ has a maximum for $N_1 = N/2$.

2. *Calculation of Δ for Large N.*

The results of the marble experiments plotted in Fig. 21 – 8 show that, at least for $N = 10$ and $N = 100$, the width Δ of the probability curve decreases and its height $NP(N/2)$ increases with increasing N, while the area under the curve remains equal to 1. The area, however, must be approximately equal to the product of width and height, for we can approximate the curve by an isosceles triangle of height $N\mathcal{P}(N/2)$ and base 2Δ. Thus the width will be approximately

$$\Delta \approx \frac{1}{N\mathcal{P}(N/2)} = \frac{2^N (N/2)!^2}{N \times N!}$$

To see what this expression means, we express $N!$ in an approximate form, known as Stirling's formula:

$$N! \approx N^N e^{-N} \sqrt{2\pi N}$$

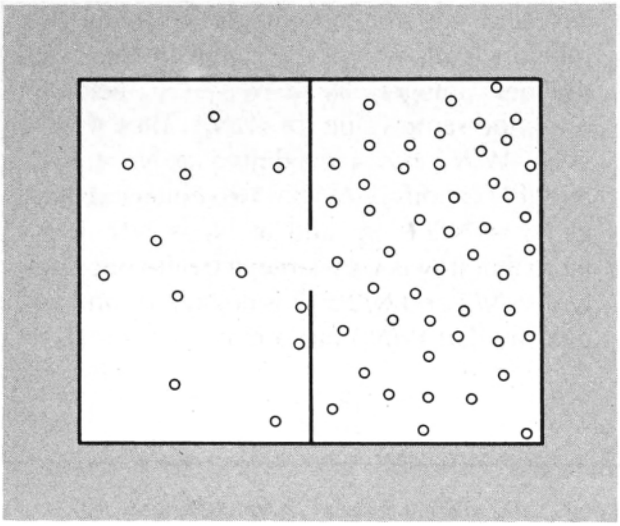
($e = 2.718 \ldots$ is a constant). It is a good approximation for large N in the sense that the ratio of $N!$ to $N^N e^{-N} \sqrt{2\pi N}$ tends to 1 as N approaches infinity. Substituting Stirling's formula in the expression for Δ we have

$$\Delta \approx \frac{2^N \left[(N/2)^{N/2} e^{-N/2} \sqrt{\pi N}\,\right]^2}{N^{N+1} e^{-N} \sqrt{2\pi N}} = \sqrt{\frac{\pi}{2N}}$$

or

$$\Delta \approx \frac{1.3}{\sqrt{N}}$$

Our discussion has been confined to the long-time behavior of the marbles under repeated shuffling, that is, the behavior after a sufficiently large number of shuffles. The short-time behavior is more complicated, because it depends on the details of the machine, such as the size of the window, and on the dynamics of the collisions of marbles. Although we cannot give a quantitative account of it using the simple ideas developed, we can understand its general features. An examination of the graphs in Fig. 21 – 2 shows that regardless of the size of the window N_1 tends to decrease toward the value $N/2$. This means that the distribution tends to evolve from a less probable one to a more probable one, in conformity with the idea that the marbles tend to be more and more randomly distributed between the two slots, as we make more and more shuffles.



21–9 A gas expanding freely into an originally empty compartment.

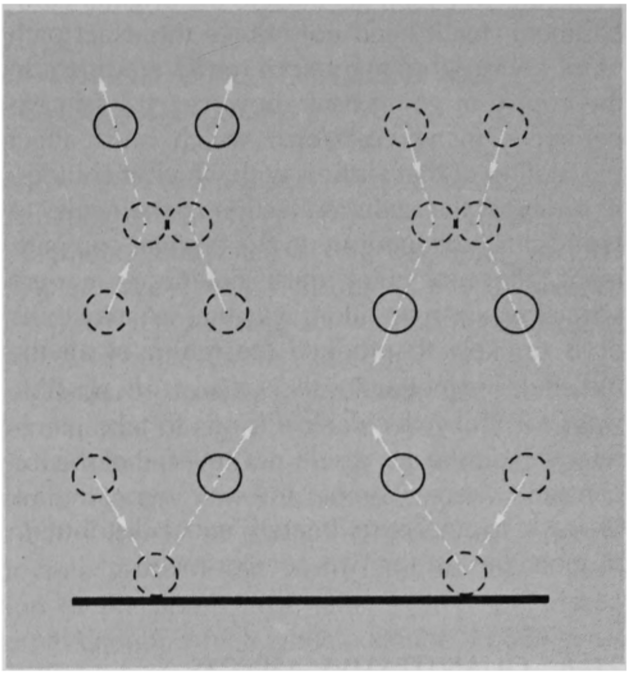
21–7 FREE EXPANSION OF A GAS

We now turn to a discussion of the free expansion of a gas into empty space, which we mentioned at the beginning of the chapter. We think of the gas as a collection of molecules confined by perfectly reflecting walls. The molecules frequently collide with the walls and with one another like perfectly elastic marbles. A schematic picture of a gas during free expansion is shown in Fig. 21 – 9. To get a feeling for the orders of magnitude of relevant numbers, let the volume of the whole container be 22.4 liters. Then we should imagine that the total number of atoms in Fig. 21 – 9 is Avogadro's number, 6×10^{23}. They fly about at an average speed of 500 meters per second for air, and each atom makes approximately 10^9 collisions per second. Since energy is conserved in the collisions, molecular motion never ceases, and the total energy of the gas remains constant.

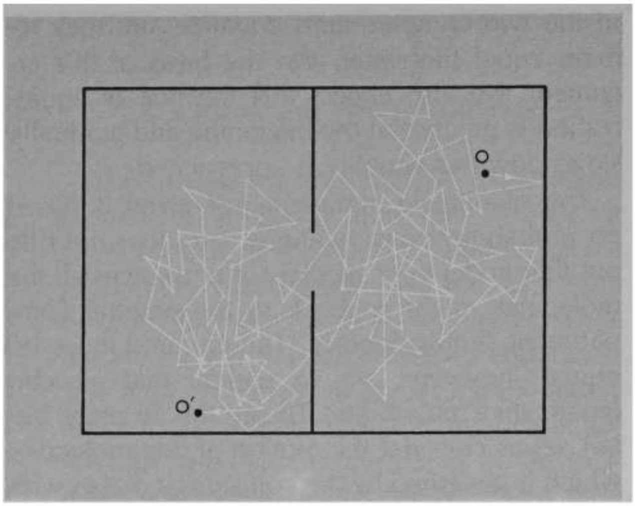
In the free expansion of a gas we are concerned only with the spatial distribution of the gas, and not with the velocities of the molecules. There is an obvious parallel between the gas under consideration and the marbles in our experiment. In our experiment we started with all the marbles in one slot and observed that they redistributed themselves equally between the two slots. In the gas we start with all the molecules in one compartment and observe that they redistribute themselves equally between the two compartments. Apart from the fact that the marble machine contains 100 marbles whereas the gas contains about 6×10^{23} atoms, they differ only in the mechanism that changes the state. In the marble experiment it was the tilting of the machine back and forth. In the gas it is collisions of atoms, either with each other or with the walls. These collisions enable an atom to bounce around the original compartment and go by chance through the window into the other compartment. Because of the close similarity between the two cases, we can make use of the intuition gained from the former to understand the latter.

Let us first consider qualitatively what happens to the molecules in the original compartment after the window is opened. After a short time, some molecules are bound to get into the other compartment, so that the number in the original compartment starts to decrease. Some of the molecules that have come from the original compartment will find their way back through the window, but there will be more molecules crossing over from the original compartment. The chance for one molecule to go from one compartment to the other is the same in both directions, but if there are more molecules in the original compartment than in the other, there are more molecules to take advantage of the chance to go one way than the other. Consequently, on the basis of pure chance, we expect a steady decrease of the number of molecules in the original compartment and a corresponding increase of the number in the other. This trend continues until finally the numbers in the two compartments equalize and they remain equal thereafter. On the basis of this argument, we also expect that the rate of equalization is greatest at the beginning and gradually slows down as equality is approached.

The qualitative argument just given is based on probability considerations, and does not rule out that under very special circumstances all the molecules might go back to the original compartment simultaneously. The circumstances required, however, are so special that we can ignore their possibility. To see this in more detail, let us consider the motion of one molecule, which is governed by the collisions it makes with other molecules and with the wall. There is no inherent irreversibility in the collision process, because if a particular collision can take place, so can its reverse, as illustrated in Fig. 21 – 10 for single collisions. Similarly, if a whole sequence of collisions is possible, the reverse sequence is also possible. When a molecule goes from the original compartment to the other, its path may



21–10 The two figures at the top illustrate the reversibility of a collision between two molecules. Those at the bottom illustrate the reversibility of the reflection of a molecule from a wall.



21–11 A possible path taken by a molecule to go from one compartment to the other. It went from the point O to the point O'. Each angle in the path is due to a collision with another molecule (not shown) or with a wall. (Not to scale.)

look like that connecting the points O and O' in Fig. 21 – 11. Each little kink in the path represents a collision with another molecule or with the wall. It is quite likely that one particular molecule will soon get back to the original compartment, for it need not retrace the exact path $O'O$, and a large number of routes are open. In the course of going back, however, it again experiences many collisions, which must affect the motion of the partners with which it collides. In order for the entire collection of molecules to get back simultaneously to the original compartment, the molecules must collide in a very special way. Any random way that we may think of is unlikely to produce the return of all the molecules simultaneously. Out of all possible ways for the molecular collisions to take place, relatively few ways result in a reversal of the expansion, whereas almost any way you can think of leads to an approximately equal distribution of molecules in the two compartments.

21–8 QUANTITATIVE ASPECTS

Let us now make a more quantitative treatment of the problem along the lines used in our experiment with marbles. Unlike the shuffling in the marble experiment, collisions in a gas do not all occur at the same instant at some signal, but they go on constantly throughout the gas. We may, however, choose to consider the state of the gas only at regular time intervals. Let us choose it to be a time characteristic of molecular collisions, the average time between two collisions experienced by the same molecule, which is about 10^{-9} sec. We thus have a time sequence of states, each evolving from its predecessor after 10^{-9} seconds. Since an average molecule has made one collision during this time, a total of about 3×10^{23} collisions have occurred throughout the gas. (Remember that it takes two to make a collision.) Between one state and its successor, therefore, the molecules are thoroughly scrambled. It is reasonable to assume that after a sufficiently long time a molecule originally in one compartment will be equally likely to be in either compartment. In terms of the state, this means that after a sufficiently long time the assumption of equal *a priori* probability holds, i.e., the state is equally likely to be any one of all the possible states. The length of time required depends on the size of the window and on the dynamics of collisions, and cannot be obtained on the basis of simple arguments.

Mathematically, our problem now reduces to the earlier problem of the marbles, because both the definition of the state and the assumption of equal *a priori* probability are the same in the two cases. We can therefore take over all the mathematical results obtained earlier.

A sufficiently long time after the initial time, the probability that N_1 molecules will be in one compartment (either one) is $\mathcal{P}(N_1)$ as given earlier, with $N = 6 \times 10^{23}$. The probability that all the molecules will be found again in the original compartment is

$$\mathcal{P}(N) = 2^{-6\times10^{23}} \approx 10^{-2\times10^{23}}$$

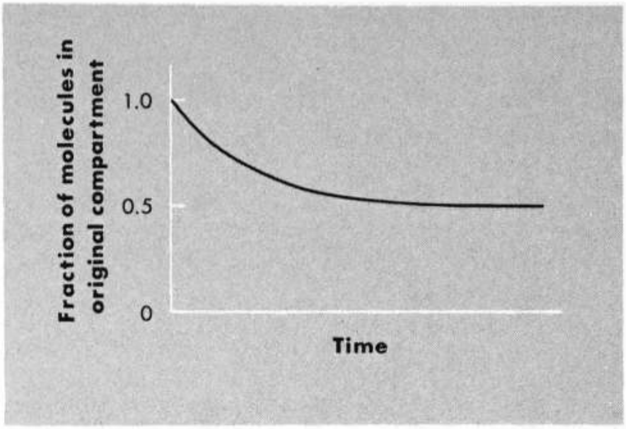
which is so small that it never happens during any physically meaningful period of time. If we were to make a graph of $N\mathcal{P}(N_1)$ against N_1/N, we would obtain a curve of the same general shape as those shown in Fig. 21 – 8, except that the peak at $N_1/N = \frac{1}{2}$ would be needle-sharp, with a width of $\Delta \approx 10^{-12}$, which is essentially zero. The behavior of the spatial distribution of molecules as a function of time is qualitatively shown in Fig. 21 – 12. It is similar to Fig. 21 – 2, except that the fluctuations are too small to be seen.

21–9 DENSITY FLUCTUATIONS

We have established that after some time the number of molecules in two compartments will equalize. This conclusion also implies that the molecules will fill the compartments with uniform density. To see this let us think of a compartment as being made up of two equal subcompartments with no partition between them —i.e., the partition is all window. We can apply the same reasoning to these two subcompartments to conclude that they have equal numbers of molecules. By successively subdividing the compartments, in our imagination, into smaller and smaller subcompartments, we arrive at the conclusion that the molecules are distributed in space with uniform density.

Of course, the process of subdividing the compartment cannot go on indefinitely because the number of molecules, though large, is finite. We must stop when we reach a subcompartment that contains only relatively few molecules. Such a minimal subcompartment, however, will be of microscopic size. For example, a subcompartment containing 10^4 molecules has a volume of only about 10^{-15} cubic centimeters. The number of molecules contained in each is not quite the same. There are fluctuations of the order of $\sqrt{10^4} = 10^2$, or 1 percent of the total. This volume of 10^{-15} cm^3, however, is too small to be seen by the unaided eye. The smallest volume that our senses can distinguish is vastly greater, say, 10^{-9} cubic centimeters, which would contain 10^7 microscopic subcompartments. In such a comparatively large volume the percentage fluctuations in the number of molecules are so small that the density is uniform as far as we can discern.

Although our senses are not sufficiently keen to detect the fluctuations in the density of molecules, there are physical phenomena which depend on them. For example, a light wave with a wavelength of a few thousand angstroms will be able to "feel" the effect of such fluctuations. When a light wave passes through the atmosphere, it can be scattered by the microscopic fluctuations in the density. The scattering is more pronounced for a light wave of shorter wavelength, because it can probe a smaller volume in the atmosphere. Thus, blue light is more sensitive to fluctuations than red light and consequently is scattered more strongly by the atmosphere. There are equally important electromagnetic factors governing the scattering of light, but the "fluctuation scattering" just mentioned is partly responsible for the blue color of the sky.



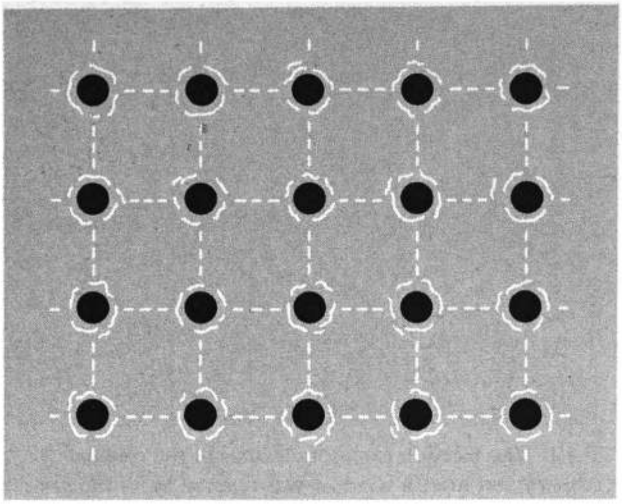
21–12 The fraction of gas molecules in the original compartment after a window was opened to let the gas expand into an adjacent compartment.

21–10 INELASTIC COLLISIONS AND HEAT CONDUCTION

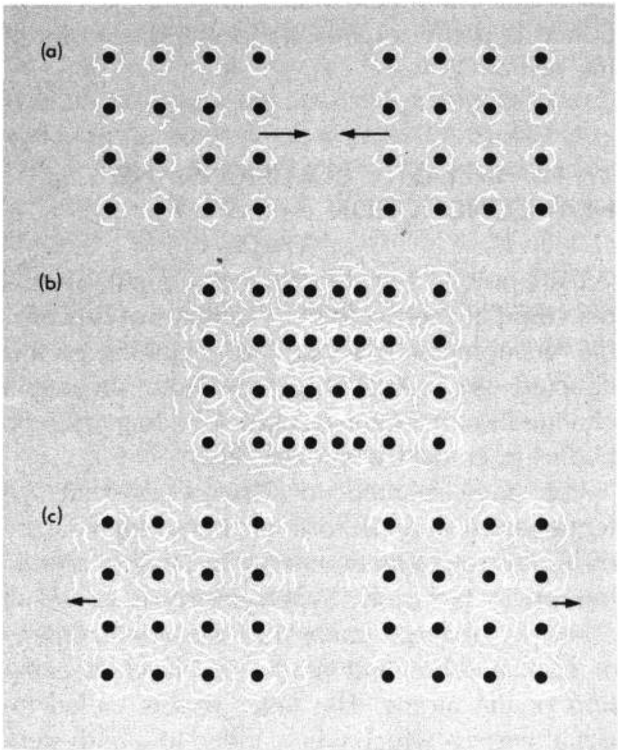
We are now in a position to give a qualitative explanation of the two other examples of irreversible phenomena cited at the beginning of this chapter. We recall that they both involve a change in temperature of two solid macroscopic bodies in contact with each other.

We have learned in Chapter 20 that the temperature of a macroscopic body is a measure of the average kinetic energy of random atomic motion in the body. When energy is added to a body, it may go partly into the kinetic energy or bulk motion, and partly into random agitation of the atoms. The latter part is called internal energy, which when added to a body generally increases the temperature.

In the example of the inelastic collision, part of the kinetic energy received by each body from the other was received in the form of internal energy, which raises the temperature of the body. We have never observed the reverse process in which one body, through spontaneous cooling, gives up internal energy to an-



21–13 **Atoms forming a crystal lattice in a solid body. They tremble irregularly about the equilibrium positions. The intensity of these oscillations is measured by the temperature. Adding heat to the body increases the trembling and correspondingly increases the temperature.**



21–14 **Magnified view of an inelastic collision between two solid bodies. In (a) they approach each other with some relative speed. When they collide, as shown in (b), the initial kinetic energy of bulk motion is partly converted into potential energy by compression of the crystal planes, and partly into random motion of the atoms. In (c) some kinetic energy is regained, but the part lost in the form of random motion is not recoverable. The solids now have higher temperature than in (a).**

other in the form of kinetic energy of bulk motion.

In the example of heat conduction, the random agitation of the atoms in one body was originally more violent than in the other. When they come into contact, the agitations tend to equalize in both bodies. The reverse process would begin with two bodies having the same degree of agitation, and upon contact the agitation in one would increase at the expense of that in the other. This process, again, we have never observed.

To understand these examples, we have to understand why the degree of random agitation of the atoms always tends to be uniform over a body. In contrast to the free expansion of a gas, we are concerned here with the problem of the distribution of energy among its various guises, instead of the distribution of atoms in space.

The atoms in a solid body are usually arranged in the form of a crystal lattice, as depicted in a two-dimensional section in Fig. 21 – 13. They are held in equilibrium positions by interatomic forces, as if they were interconnected by mattress springs. The random oscillations they make about the equilibrium positions are symbolically shown in Fig. 21 – 13 by means of the trembling lines about each atom. The temperature of the body is a measure of the violence of these motions. In fact, the trembling can become so violent that atoms break away from the crystal, corresponding to the fact that the solid melts at a sufficiently high temperature.

When a solid body moves in uniform motion, the entire lattice of Fig. 21 – 13 moves without affecting the atomic oscillations. The velocity of each atom is then composed of two parts, a part which is the translational velocity of the whole lattice, and a fluctuating part due to the random oscillations about the moving equilibrium position. The former gives rise to the kinetic energy of bulk motion, while the latter shows up as the temperature of the body.

Consider what happens when two macroscopic solids collide with each other. As shown in the sequence of pictures (a), (b), and (c) in Fig. 21 – 14, the initial kinetic energy of bulk motion in (a) is mainly converted into potential energy stored in the compressed lattice planes in (b), but a part of it agitates the atoms into more violent oscillations. When the lattice planes spring back to their natural positions, as in (c), the two bodies bounce off with regained kinetic energy, which is always smaller than the original kinetic energy, because part of that has gone into raising the temperatures of the bodies. For the reverse process to happen, we must imagine that during phase (b) of the collision the atoms in each body just happen to oscillate in unison, in the right way, at the right moment, to make the two bodies bounce back with more total kinetic energy than before the collision. Although such a process does not violate the conservation of energy and momentum, the chance that it will occur is nil, because it requires too high a degree of coincidence in the random oscillations of the atoms. In other words, the reverse process is highly unlikely, because out of the almost infinite number of ways in which the atoms can oscillate, only a few will lead to the reverse process, whereas all others will lead to what we normally observe. A quantitative formulation of this idea can be made in a way analogous to what we have done for the free expansion of a gas. The present case, however, is more complicated because the quantitative problem of energy distribution involves a detailed knowledge of the mechanics of a crystal lattice.

The irreversibility of heat conduction can also be qualitatively explained along similar lines. Here we are concerned with two lattices in contact without any compression of the lattice planes, but with the atoms in one lattice oscillating more violently than those in the other.

A detailed study of processes in which the distribution of energy among many molecules is important is the subject of the next chapter.

QUESTIONS

1. You are shown a movie of an egg being beaten by an electric beater. The movie starts when the yolk and the white are already thoroughly mixed. Could you tell whether the movie is being run forward or backward?

2.* Figure 21 – 3 shows that if you start with 100 percent of the marbles in slot 1, then after 8 shuffles the fraction has dropped to 50 percent. If you start with 50 percent of the marbles in slot 1, what fraction do you expect to find in slot 1 after 8 shuffles?

3.* Suppose 50 red marbles are placed in slot 1 and 50 blue marbles are placed in slot 2. The board is then shuffled 10 times.

(a) How many marbles do you expect to find in slot 1?

(b) How many red marbles do you expect to find in slot 1?

4.* A single coin is tossed two times. List all the possible sequences that might be observed, using H for heads and T for tails. In what fraction of the cases will the same side appear both times?

5.* Arnold has just thrown HHHH with a coin. What is the probability he will now throw an H again?

6.* What is the probability, when you roll a die, of getting (a) a "one" in a single throw, (b) a "two" and a "five" (in either order) in two throws?

7. If you draw a card from a deck of playing cards, what is the probability that it will be (a) a red card? (b) a club? (c) the ten of diamonds?

8. Two long sticks of triangular cross section are both painted so that one side is red, one blue, and the other green. If these are dropped together, what is the probability that red will show?

9.* In Fig. 21 – 4, imagine that marbles 1 and 7 were interchanged in *A* to form a new arrangement *A'*.

(a) Does *A'* belong to the same state as *A*?

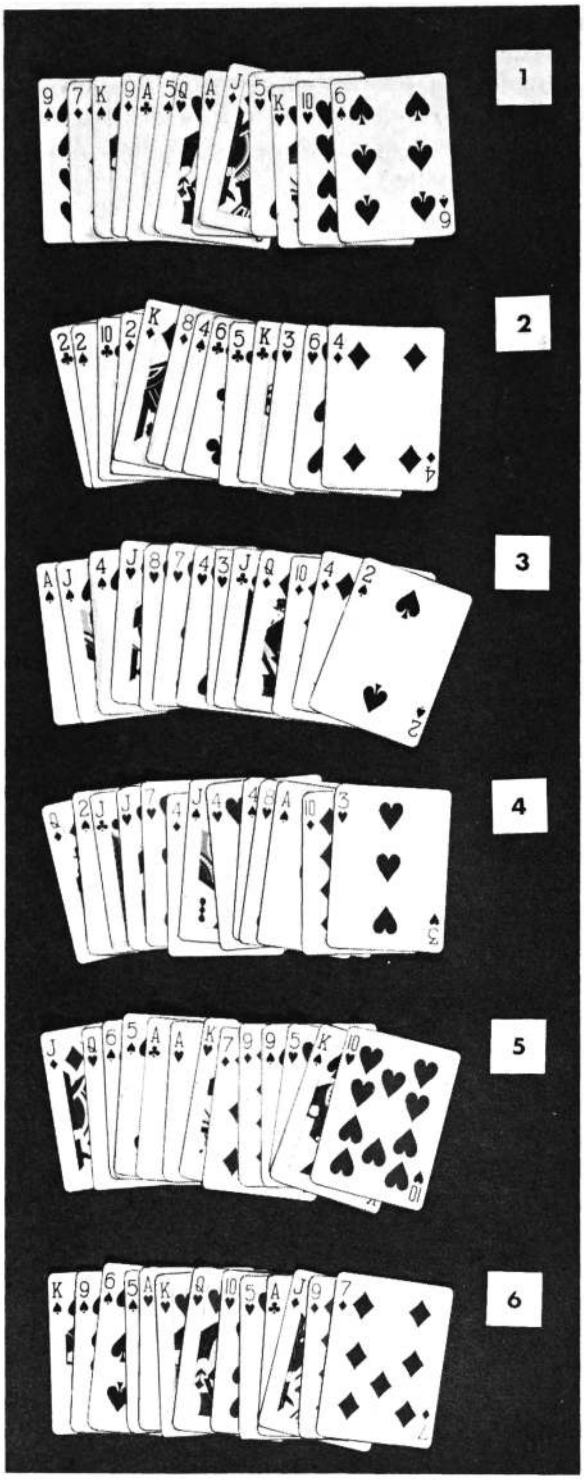
(b) Does *A'* belong to the same state as *B*?

(c) Does *A'* belong to the same state as *C*?

10.* In the PSSC film "Random Events," Dr. Hume and Dr. Ivy use a machine consisting of 16 squares, each of which has a white face and a black face.

(a) How could you specify a *state* of this system?

(b) How could you specify a *distribution* of this system?



21–15 For Question 11.

11. In Section 21 – 5, three different expressions are used in connection with the marble experiment: arrangement, state, and distribution. Figure 21 – 15 shows six hands of bridge. Identify cards with marbles and suits with slots.

 (a) How many different arrangements are there?

 (b) How many states?

 (c) How many distributions?

12.* A thought experiment: suppose you have 10,000 identical marble machines fixed side by side on the same board. Tilting the board, you tilt all the machines at the same time.

 Each machine contains 10 marbles. You tilt back and forth a certain number of times to be sure that any existing memory has been erased, and then allow the marbles to settle down in the slots.

 In how many machines do you expect to find an equal distribution of marbles?

13.* Let $\mathcal{P}(N_1)$ be the probability that N_1 marbles are in slot 1, in an experiment with N marbles. Thus N_1 can assume the values $0, 1, \ldots, N$.

 (a) What is the meaning of this expression?

$$\sum_{N_1=0}^{N} \mathcal{P}(N_1)$$

 (b) Can you state its value without calculating it?

14. What is the total number of states belonging to the distribution shown in Fig. 21 – 4?

15. Prove that with N marbles in two slots the total number of states is 2^N.

16. The area under each curve of Fig. 21 – 8 is said to be unity. Approximate the area under each of the curves by two right triangles drawn so that the extra area they include equals the area they leave out. In each case, how close to unity is the area of the two triangles?

17. (a) Expand the quantity $(p+q)^N$ and show in detail that the coefficient of the term $p^{N_1}q^{N-N_1}$ in the expansion equals the function $W(N_1)$ defined in Section 21 – 6.

 (b) By choosing special values for p and q, use the expansion in (a) to prove that $W(0) + W(1) + W(2) + \cdots + W(N) = 2^N$.

18. A marble machine contains N marbles. Let

$W(N_1)$ represent the number of states in which there are N_1 marbles in slot 1, and $W(N - N_1)$ represent the number of states in which there are $(N - N_1)$ marbles in the same slot 1.

(a) What is the relation between $W(N_1)$ and $W(N - N_1)$?

(b) Would your answer to (a) be modified if the machine showed a bias toward, say, slot 1?

19. From the data in Fig. 21 – 6 calculate the average fraction of marbles in slot 1 after the initial distribution of the 10 marbles is "forgotten." Subdivide the data into four equal groups and calculate the average separately for each group. What conclusions can you draw from your results?

20. It is sometimes said that a monkey sitting at a typewriter and typing away at random will eventually produce a complete manuscript of Shakespeare's "Hamlet." Estimate how long it would take on the average before the monkey produced merely the phrase, "To be or not to be."

21. There are about 6000 stars visible to the naked eye. Assuming that they are uniformly distributed on the average, into how many imaginary subcompartments should we divide the sky if we want to be reasonably sure that the number of stars in any two subcompartments do not differ by more than 10 percent?

22.* A biologist divides a circular sample dish of radius r into two regions by drawing a circle of radius $r/2$ about the center. He then pours in a little water and in the very center places a sample of small free-swimming organisms. What fraction of the sample will be found in the outer ring after some time?

23. An aquarium contains 15 fish of each of two species, separated from each other in two equal compartments. The fish swim around randomly and independently in such a way that, when the partition is removed, one fish crosses from one compartment into the other every second on the average. How long would it take on the average (a) for all the fish of one species to come back into their original compartment? (b) for the two species to separate, one species in one compartment and the other species in the other compartment?

24. In the free expansion of a gas from one compartment into another, what data would completely specify the state and the distribution of the gas at a given instant?

25. 6×10^{23} molecules of a gas is held in a container. Suppose you could count the number of molecules in each half of the container at any instant. By how much would you expect this number to vary from one count to another?

26. Two connected compartments of equal volume contain air at atmospheric pressure. The total number of molecules in the two compartments is 10^{24}. If the gas in the compartments is pumped out with a very good vacuum pump, it is possible to reduce the pressure to 10^{-12} atmospheres. Compare the width Δ of the curve of $N\mathcal{P}(N_1)$ vs N_1/N for these two conditions.

27. A typical living room has a volume $V = 100\ \text{m}^3$. We want to estimate the probability that a small region of the room of volume $v = 1\ \text{cm}^3$ should suddenly, spontaneously, empty itself of all air molecules. Let N be the total number of air molecules in the room. State your assumptions explicitly.

(a) What is the probability that a particular air molecule in the room should be in the small volume v at a given instant?

(b) What is the probability that it is not in the volume v?

(c) What is the probability that none of the N molecules in the room should be in the volume v? Give your answer in terms of v, V, and N.

(d) Estimate N and calculate the numerical answer to (c). Use the following mathematical facts: For any positive number $x \ll 1$, $(1 + x)^k = 10^{k \log(1+x)} = 10^{k \ln(1+x) \log e} \approx 10^{kx \log e} \approx 10^{0.4kx}$.

(e) What is this probability if v is a cube of side one angstrom?

22 Entropy

22–1 THE MACROSCOPIC STATE

IN THE PRECEDING chapter we saw several examples of processes which proceed spontaneously in one direction but not in the other. To see what is at the root of such processes we analyzed an experiment with marbles. We have demonstrated that whatever the initial distribution of the one hundred marbles between the two slots, after a sufficient number of tiltings the distribution remained almost even. We reasoned that while all states of the marbles are equiprobable, the even distribution corresponds to the largest number of such states and is, therefore, the most probable distribution. Whereas specifying a state requires the detailed knowledge of which marbles are in which slot, specifying a distribution requires only the knowledge of the number of marbles in each slot.

We used the same language to describe the free expansion of a gas from one compartment to another, but here a description in terms of the pressure or the density of the gas in each compartment is more convenient, since these quantities are directly measurable. Similarly, when we come to analyze such irreversible processes as an inelastic collision between two putty balls, we describe the system in terms of directly measurable quantities such as temperature and bulk kinetic energy. We call the description of a system in terms of such directly measurable properties the *macroscopic* state, to distinguish it from the *microscopic* state (called "state" for short in the preceding chapter), which is the description requiring a detailed knowledge of the position and energy of each atom.

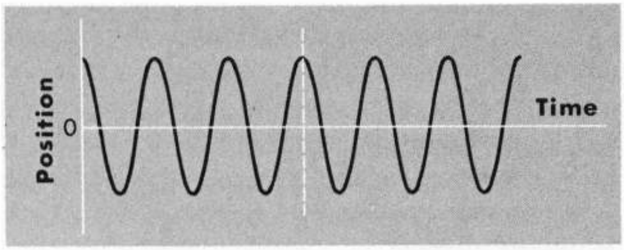
Our task now is to find a measure to compare the probabilities of different macroscopic states of a system. For example, how much more likely is it that two blocks of metal in thermal contact will remain at the same temperature than that one of them will warm up spontaneously while the other cools off?

To do this we shall first study reversible processes, that is, processes which can go in either direction. In such processes the macroscopic states of the system may be very different from one another, but the fact that the process is reversible suggests that all the macroscopic states involved have substantially the same probability.

22–2 REVERSIBLE PROCESSES

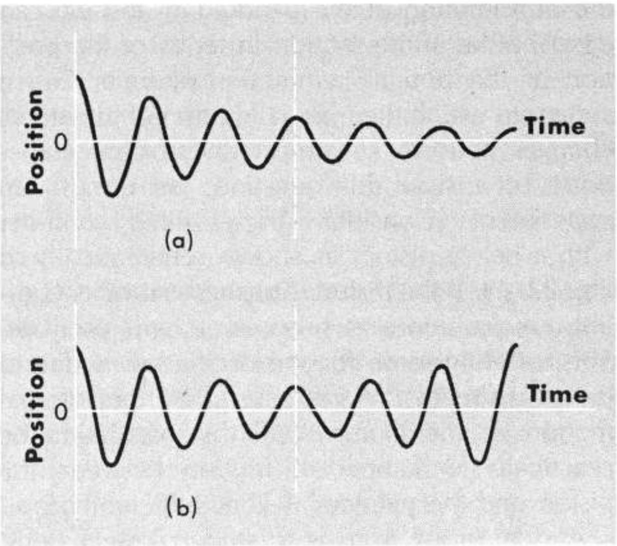
A system consisting of many atoms, say of the order of a mole,* can undergo processes that are almost reversible as long as the internal motion of the atoms remains practically unaffected. Planetary motion provides an illustration of a nearly reversible process. You cannot tell by looking at a motion picture showing the motion of the earth in its orbit whether the film is running forward or backward. As another example, consider a full oscillation of a pendulum starting from the left and swinging first to the right and then to the left. If it is a pendulum with practically frictionless suspension and a heavy bob to minimize air resistance, the swing to the left will be very nearly as high as the swing to the right. You could not tell by looking at a motion picture whether you were seeing the real or inverse projection; the oscillation might just as well have started with a swing to the left.

If we let the pendulum make many oscillations and plot the position of the bob as a function of time, we shall get a curve which, to all practical purposes, will be indistinguishable from a sinusoidal curve (Fig. 22 – 1). If from the instant shown by the dashed line in Fig. 22 – 1 we had plotted the position of the bob as a function of time up to that moment as it would be seen in a motion picture which is screened backward, the curve would coincide with the one that describes the actual motion. A heavily damped oscillation, on the other hand, is clearly irreversible. A plot of the position of the bob versus time might look like the curve shown in Fig. 22 – 2 (a). If we replace the part of the curve to the right of the dashed line by the curve describing the motion up to that moment as it would be seen on a backward-screened film, we would get Fig. 22 – 2(b), which does not correspond to any free oscillation that has ever been observed.



22–1 **The position-time plot of a pendulum bob swinging from a practically frictionless suspension. The oscillations to the left of the dashed line have been replotted to the right so that they correspond to what would be seen in a movie run backward. Since there is no difference in the two sections of the graph, we say that the motion is reversible.**

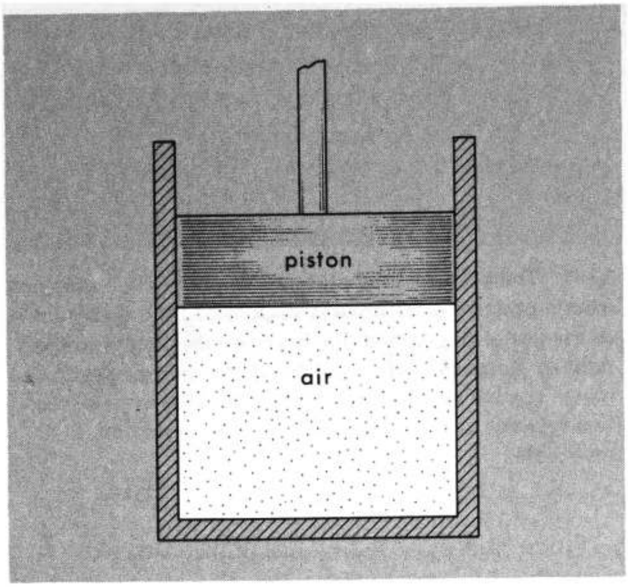
Since there is really no pendulum which is completely free of friction, the motion of all pendulums is, in the long run, damped and therefore irreversible. A convenient measure for the degree of irreversibility in the motion of the pendulum is the time it takes for the amplitude of the oscillation to be reduced to half its initial value. We shall refer to this time as the *damping time*. An undamped oscillation is then the idealized limit of a damped oscillation with an infinite damping time. Since we shall be interested in extrapolating from the real to the ideal case, it will be convenient to use the



22–2 **If the pendulum oscillations are heavily damped, the plot of the actual motion to the right of the dashed line in (a) does not look like the plot to the right of the dashed line in (b), which corresponds to a movie of the first part of the motion run backward. In this case, the motion is irreversible.**

---

*A mole (6.02×10^{23}) is a convenient way of expressing large numbers such as those involved in various atomic phenomena. We shall use it occasionally from now on.



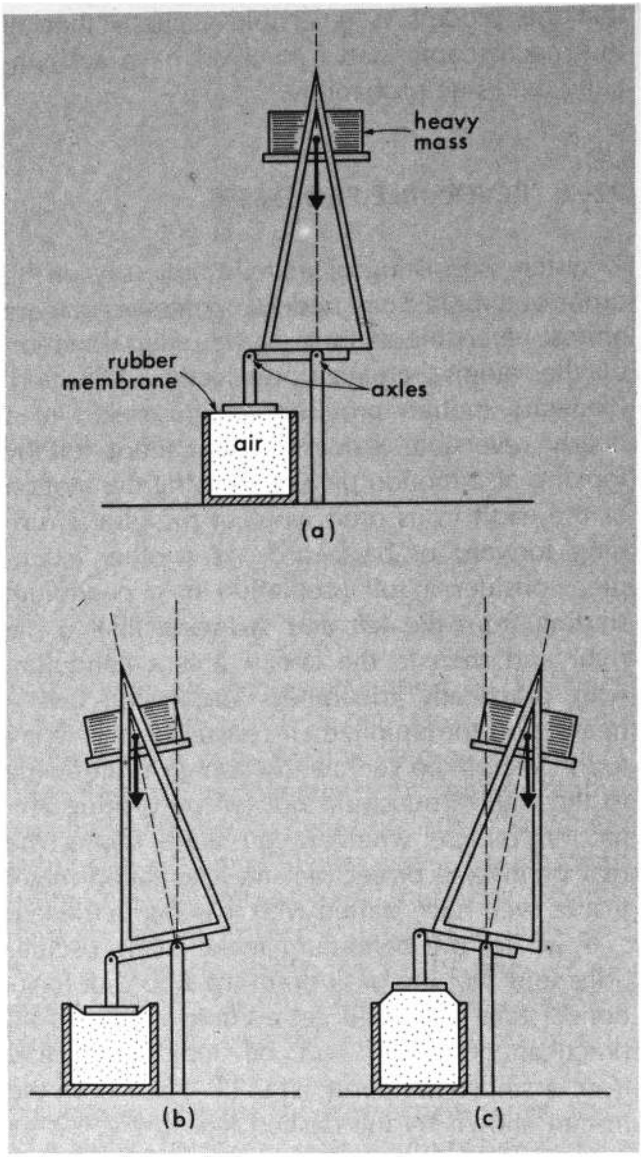
22–3 **If the piston is lowered or raised, a restoring force due to the change in air pressure will cause it to oscillate around its equilibrium position.**

inverse of the damping time, which is called the *damping coefficient*. Thus, an undamped oscillation corresponds to a damping coefficient equal to zero.

22–3 THE GAS OSCILLATOR

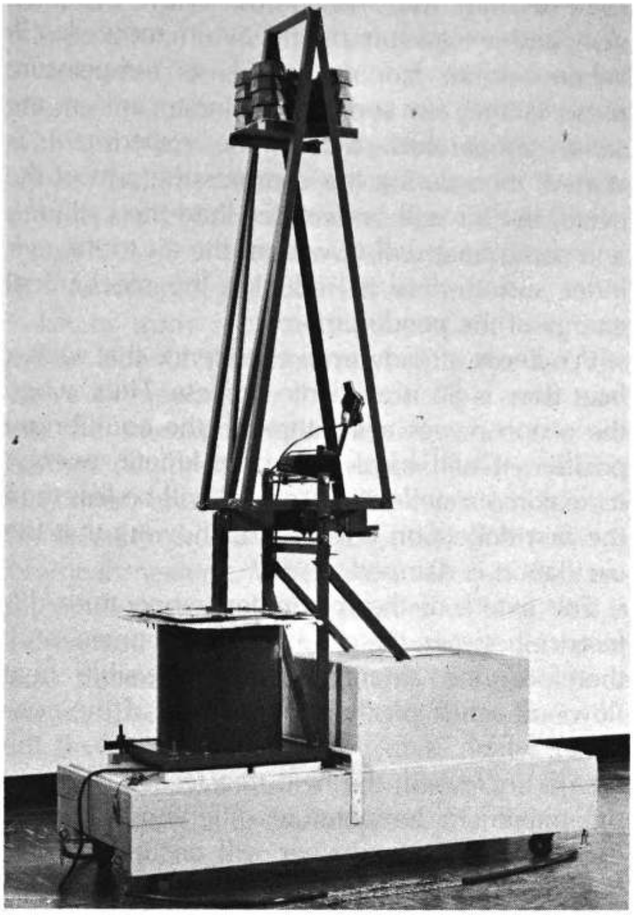
In the ideal, frictionless case the internal motion of the atoms of the pendulum is not affected by the bulk motion of the pendulum, and we can describe the entire motion in terms of the position of the bob as a function of time. Can a system in which the internal motion of the atoms changes perform an almost undamped oscillation? To answer this question, we turn to an experiment. Consider an air-filled cylinder with a heavy piston, as shown schematically in Fig. 22 – 3. If the piston is pushed out of its equilibrium position and released, it will oscillate. The restoring force for the oscillation is due to the pressure of the gas caused by the internal motion of the atoms. Will this oscillation be practically undamped if friction between the piston and the cylinder is kept sufficiently low?

As you know from your study of mechanics, it takes special care to minimize piston friction and yet have no leakage of air from the cylinder. To achieve this, we connect the piston to the cylinder with an elastic rubber membrane which has very little internal friction and a very small restoring force. Thus, if the piston is pushed out of the equilibrium position, it is the force due to the change in gas pressure inside the cylinder which accelerates it, and not the tension in the rubber membrane.



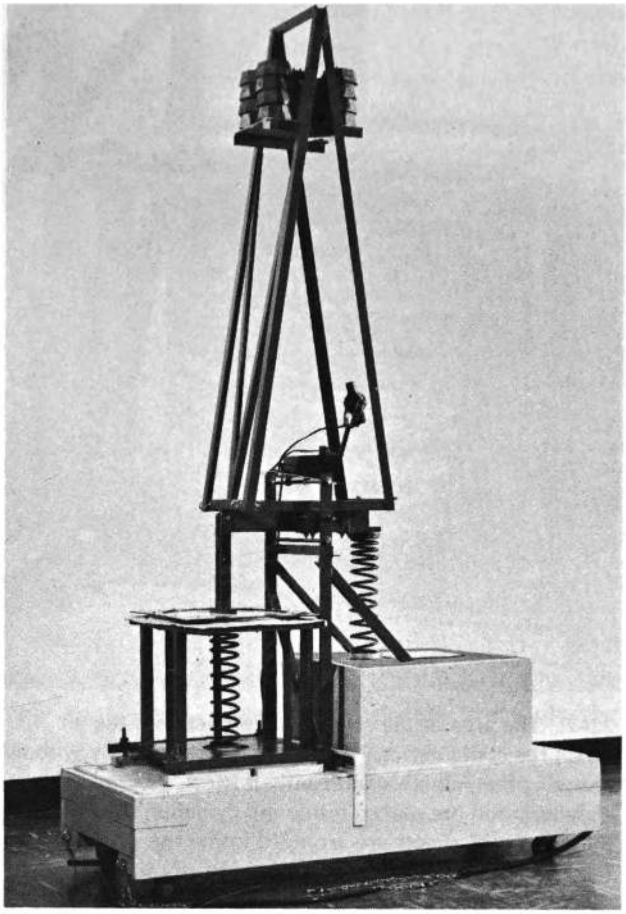
22–4 **(a) The inverted pendulum consists of a rigid structure whose mass is concentrated near its top and which is free to rotate about an axle passing through fixed bearings supporting it from below. The inverted pendulum is attached through another axle to the piston of the air-filled cylinder. (b) and (c) As the piston moves down and up, the inverted pendulum swings to the left and right. An elastic rubber membrane between the cylinder and the rim of the piston seals the air in the cylinder without adding any significant friction to the system.**

When the gas in the cylinder is compressed, its pressure and temperature rise; when the gas expands, its pressure and temperature fall. If we want to be sure that at each instant both the pressure and the temperature are the same throughout the system, the motion of the piston



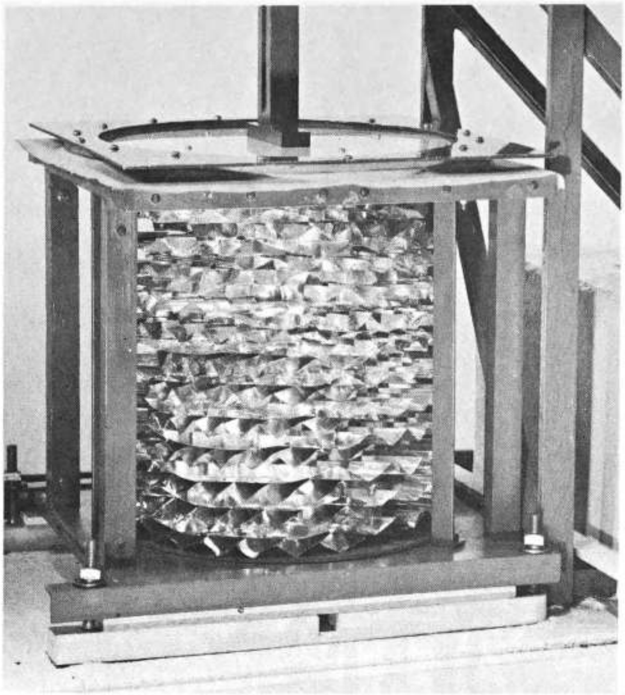
22–5 A gas oscillator constructed as shown in Fig. 22–4. A fixed light source is mounted on the fixed central support just above and to the right of the central axle. This light source casts a narrow beam of light on a mirror mounted on the underside of the platform holding the lead masses at the top of the pendulum. A thin wire (not visible in the figure) mounted near the lower right-hand side of the pendulum is in the reflected light beam. As the pendulum oscillates, the shadow of this wire moves back and forth across a horizontal sheet of white paper placed on top of the cement block to the right of the cylinder, indicating the amplitude of the oscillations.

must be very slow at all times. This could be achieved by making the piston very heavy, but it would be rather cumbersome. We can produce the same effect while keeping the oscillating mass within manageable limits by applying a force in the opposite direction to the restoring force of the gas. This can be done by an inverted pendulum (Fig. 22 – 4). When the piston is in the equilibrium position (a), the weight on the top of the pendulum acts only on its bearings. When the air is compressed and therefore pushes up on the piston (b), the pendulum is off balance and part of its weight pushes down on the piston. When the piston is above the equilibrium position (c), the air pressure inside the cylinder is less than that outside, and therefore there is a force pushing the piston down. But now the pendulum leans to the other side, and therefore its weight tends to pull the piston up, again reducing the restoring force. An example of this kind of oscillator is shown in Fig. 22 – 5. All the data discussed in this section were taken with this instrument.

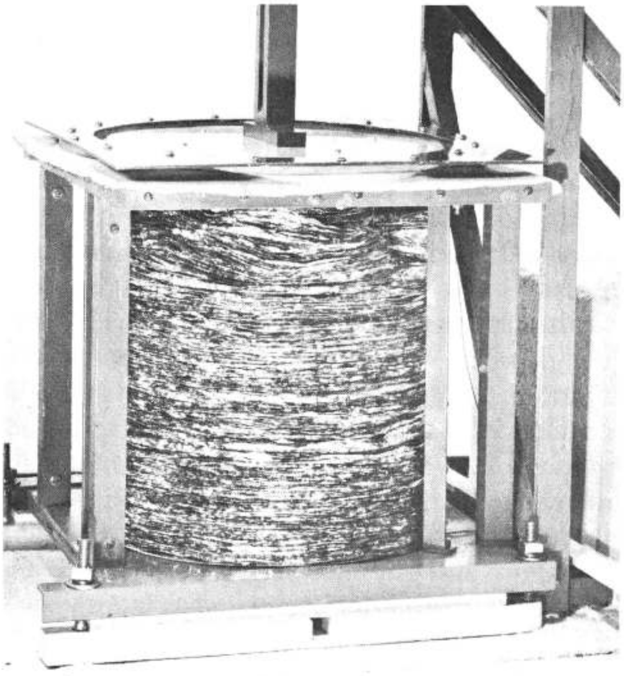


22–6 When the cylinder and piston are replaced by two steel springs, the effect of bearing and air friction alone can be measured.

The first measurement was a determination of the damping time. As you might expect, it is finite. In fact, it was found to be 7.6 cycles of oscillation. This corresponds to a damping coefficient of 0.13 $(\text{cycles})^{-1}$. One might be inclined to blame the mechanical friction for all the damping. But a direct check was made on the effect of the friction in the bearings and the air resistance (outside the cylinder) by replacing the cylinder with two steel springs which produced the same period of oscillation as the gas (Fig. 22 – 6). This revealed that bearing friction and air resistance account for only 0.04 $(\text{cycles})^{-1}$. Where does the other 0.09 $(\text{cycles})^{-1}$ come from?



22–7 **The area of the surface in contact with the air inside the container can be increased considerably without changing the volume significantly if sheets of crinkled aluminum foil are placed inside the cylinder. The figure shows how 25 sheets are arranged inside the cylinder. The cylinder has been removed temporarily so that the sheets would be visible in the photographs.**



22–8 **The appearance of the sheets when 400 of them are stacked in the cylinder.**

Let us look into the process of the compression and expansion of the air in more detail. When the air is compressed, its temperature rises. If the air and the cylinder are at the same temperature before the experiment is started, then during the compression part of the cycle, the air will be warmer than the cylinder, and some heat will flow from the air to the cylinder surrounding it, reducing the mechanical energy of the pendulum.

We know already from experience that such a heat flow is an irreversible process. Thus, when the piston passes again through the equilibrium position it will have lost some kinetic energy. Its maximum deflection upward will be less than the first deflection downward, showing that the oscillation is damped.

The extent of the maximum temperature difference between the gas and the container—and, therefore, the extent of the irreversible heat flow—depends on the surface area of the container which is in contact with the gas. If the area is increased, the heat transfer is faster and the maximum temperature difference between the gas and the container will decrease. Thus, the temperature changes within the gas during the oscillation also will decrease. How will such a decrease affect the damping?

A convenient way of increasing the surface area of the container with only a negligible change in the volume of the gas and without any change in the moving parts of the oscillator is to introduce a large number of thin aluminum sheets into the cylinder. The arrangement of sheets is shown in Figs. 22 – 7 and 22 – 8. The sheets of aluminum, together with the cylinder and the piston, have a very large heat capacity; that is, they can absorb a large quantity of heat with practically no change in temperature.

The maximum temperature change of the gas for different numbers of aluminum sheets, and the corresponding damping time and damping coefficients, are given in Table 1. The temperature of the gas was measured by the use of a resistance thermometer. The initial amplitude was the same in all cases. The total surface area of the cylinder and the piston was equal in area to 3.5 sheets. The damping coefficient due to mechanical friction was found to be about 0.04 $(\text{cycles})^{-1}$, independent of the number of aluminum sheets. We have subtracted this value from the value computed from the observed damping time to give the last column in Table 1.

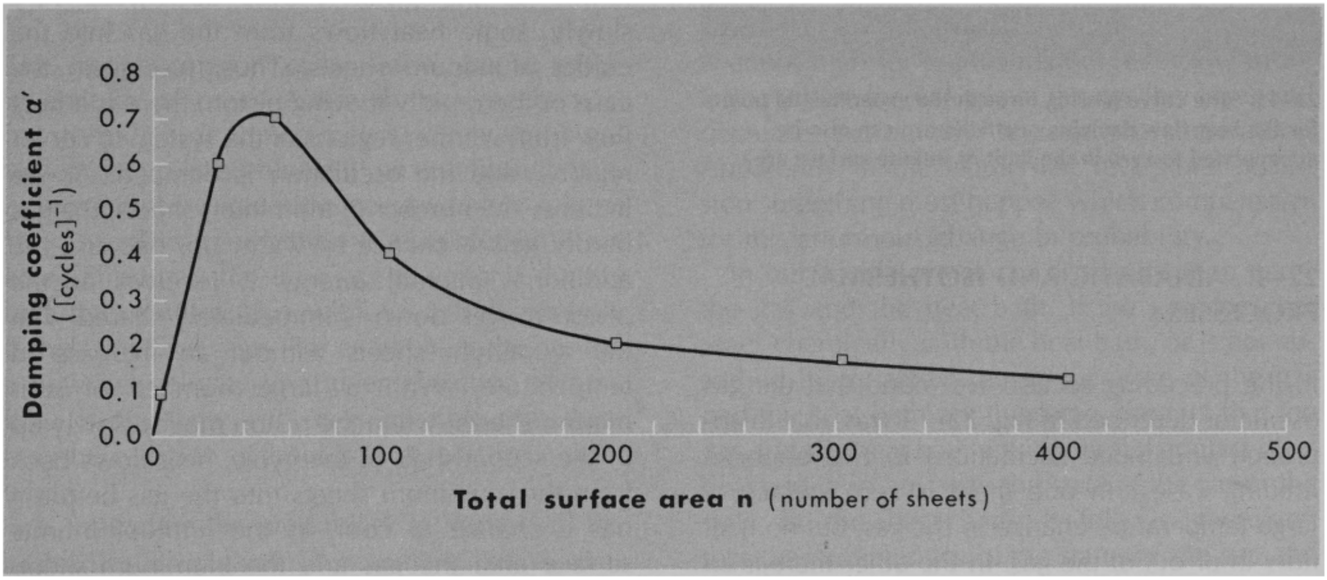
We see from Table 1 that the maximum change in the temperature of the gas decreases with the increase in the number of aluminum sheets, but the damping coefficient α' first increases sharply and then decreases rather slowly. Figure 22 – 9 shows the damping coefficient α' as a function of the total surface area n.

Let us study separately parts of the curve on either side of the maximum. On the left, the curve extrapolates to the origin. This point cannot be checked experimentally, because we cannot compress a gas without having it in a container, and the walls of a container have a finite surface. Nevertheless, the point has a physical meaning. When the surface of the container is decreased, the amount of heat that is exchanged between the air and the container also decreases. Thus the limit of zero surface means that no heat is exchanged between the gas and the container. In this case, when the air is compressed, some of the mechanical energy of the system is converted into internal energy of the air without an increase in the internal energy of the container. When the gas expands, the same amount of internal energy is converted back into mechanical energy. If we could eliminate friction entirely, the oscillations would not be damped; the process would be reversible.

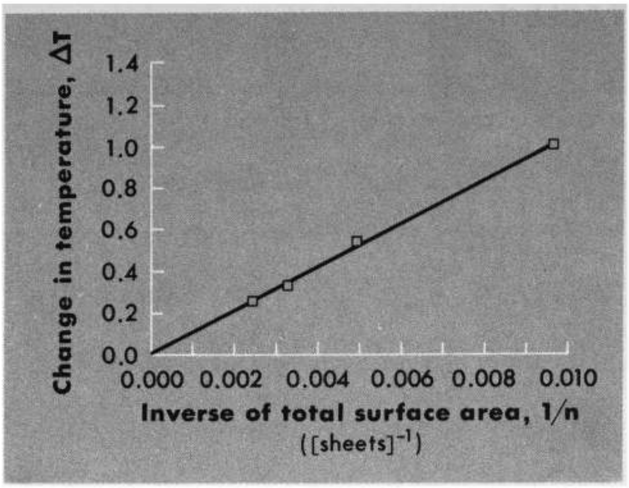
TABLE 1

SURFACE AREA n IN SHEETS	TEMPERATURE CHANGE ΔT IN ARBITRARY UNITS	DAMPING TIME t_d IN CYCLES	DAMPING COEFFICIENT IN (CYCLES)$^{-1}$ α $(1/t_d)$	α' $(\alpha - 0.04)$
3.5	4.1	7.6	0.13	0.09
28.5	—	1.57	0.64	0.60
53.5	—	1.35	0.74	0.70
103.5	1.00	2.3	0.44	0.40
203.5	0.54	4.1	0.24	0.20
303.5	0.33	5.0	0.20	0.16
403.5	0.25	6.3	0.16	0.12

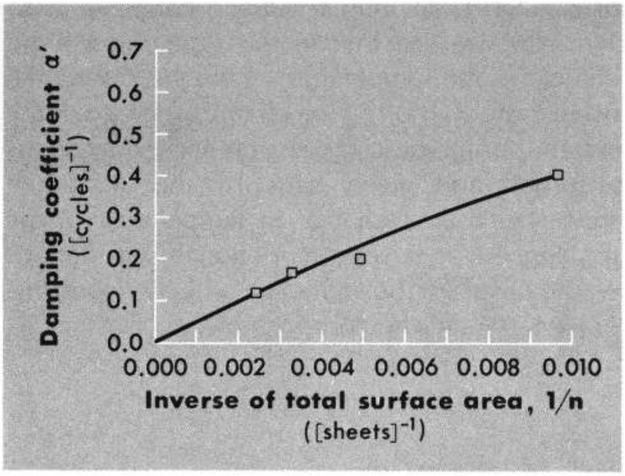
To extrapolate the data in Table 1 to an infinite surface area, we plot both ΔT and α' as a function of $1/n$ (Figs. 22 – 10 and 22 – 11). We see that both graphs extrapolate through the origin. Since an infinite surface corresponds to $1/n = 0$, we see that in this limit there is no change in the temperature of the gas during the oscillation. Part of the mechanical energy of the system during the compression and expansion is absorbed and given back by the aluminum sheets without a change in temperature of the gas. Since $\alpha' = 0$, without friction the oscillations would also not be damped, and therefore this process would also be reversible.



22–9 The damping coefficient due to heat flow as a function of surface area, of the gas oscillator shown in Fig. 22–5.



22–10 The change in temperature of the enclosed air during one cycle of oscillation plotted against the inverse of the surface area. The curve through the experimental points can be extrapolated to zero in the limit of infinite surface area. This means that in the limit, the air temperature remains constant.



22–11 The curve passing through the experimental points for the heat flow damping coefficient α' can also be extrapolated to zero in the limit of infinite surface area.

22–4 ADIABATIC AND ISOTHERMAL PROCESSES

In the preceding section we found that the gas oscillator described in Fig. 22 – 5 was able to approach undamped oscillations in two opposite limiting cases: In one there is zero surface, a large temperature change in the gas, but no heat flow in or out of the gas. In the other there is an infinite surface, no temperature change in the gas, but a large heat flow. Let us study these cases in more detail.

In the first case no heat is exchanged with the container. Processes where no heat flows in or out of the system are called *adiabatic* processes. It is natural to ask now whether all adiabatic processes are necessarily reversible. Consider for example, an oscillator like that in Fig. 22 – 5, but with a considerably stronger restoring force. Suppose we raise the piston and release it. As the piston rushes downward it compresses the air near it and in so doing raises its temperature. Turbulences may occur, and for a while the air near the piston will be warmer than the air farther down the cylinder. Heat will flow from the warmer region to the cooler one, tending to equalize the temperatures. Thus the air farther down the cylinder acts like the walls of an imaginary container. The oscillation will be damped even in the limit where no heat flows through the surface of the actual container. There will be no heat flow from a warmer to a colder region only if the piston moves slowly enough so that the rise in temperature occurs throughout the air almost simultaneously. We can expect, therefore, that the only adiabatic processes which are reversible are those which take place so slowly that at any instant the temperature is the same throughout the system.

Let us now consider our oscillator with, say, 50 sheets of aluminum in the cylinder, and again analyze what happens when we raise the piston and release it. Although the gas warms up uniformly as the piston moves downward very slowly, some heat flows from the gas into the colder aluminum sheets. Thus, just as in the case of the rapidly moving piston, there is a heat flow from warmer regions of the system to cooler regions, and the oscillation is damped. As we increase the number of aluminum sheets, the gas hardly gets a chance to warm up, because the additional internal energy it receives as the piston moves down is immediately shared with the aluminum sheets without an increase in temperature. With a large number of aluminum sheets, when the piston moves slowly up in the second half of the cycle, heat flows back from the aluminum sheets into the gas before it has a chance to cool. In the limit of infinite surface area the gas and the aluminum sheets stay at the same temperature, heat flows equally well in either direction, and part of the mechanical energy of the piston can be converted into internal energy in the aluminum sheets and back again. Since there is no change in temperature during the entire cycle, the process is called *isothermal.*

A word of caution may be in place here: an isothermal process is not necessarily reversible. In Chapter 21 we studied the free expansion of an ideal gas: When an ideal gas leaks from one container into another container which was previously evacuated, there is no change in the temperature of the gas. But the process is not reversible, even if the gas leaks very slowly through a tiny hole. Although the bulk of the gas is at rest in both containers, and the temperature does not change, the gas near the hole has a net velocity and is therefore substantially in a different condition from the rest. Apparently, for an isothermal process to be reversible the same conditions must prevail throughout the gas. Not only must the reversible isothermal oscillation be slow; it requires good thermal contact with a body of very large heat capacity to ensure the same conditions throughout the system. There is nothing special about aluminum sheets in this respect. We could have used a large number of thin copper strips evenly distributed. We shall refer to any such arrangement which provides a large heat capacity and good thermal contact as a heat bath.

22–5 FREE EXPANSION AND ISOTHERMAL EXPANSION OF AN IDEAL GAS

The free expansion of an ideal gas is an irreversible process. Although no mechanical law forbids it, it is extremely unlikely that a gas which occupies a volume V_2 at equilibrium will, as a result of a fluctuation, occupy only a volume $V_1 < V_2$. This observation is restricted to a gas at equilibrium, that is, a gas at uniform density and temperature and without a bulk motion. If this is not the case—for example, if the gas has a high net velocity in a given direction and the flow is suddenly cut—it is possible that there may be a region without gas, at least for a short while.

To calculate the probability of finding all the molecules of a gas in a volume V_1 or V_2, it is convenient to use a reference volume V_0 which contains all the molecules of the gas and which is larger than either V_1 or V_2. If the gas is in equilibrium, a molecule is equally likely to be in any particular part of the volume V_0. Since the probability of finding it in the volume V_0 is 1, the probability of finding it in V_1 is V_1/V_0. In an ideal gas the location of one molecule is independent of the location of the others. Thus, the probability of finding two molecules in the volume V_1 is $(V_1/V_0) \cdot (V_1/V_0) = (V_1/V_0)^2$ and to find all N molecules in V_1 the probability is

$$\mathcal{P}_1 = (V_1/V_0)^N$$

Similarly, the probability of finding all the gas in a volume V_2 is

$$\mathcal{P}_2 = (V_2/V_0)^N$$

The ratio of these probabilities

$$\mathcal{P}_1/\mathcal{P}_2 = (V_1/V_2)^N$$

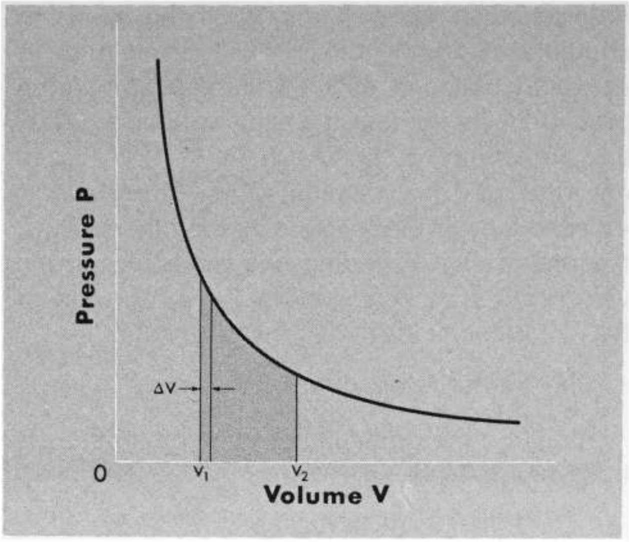
is independent of the reference volume. When $V_1/V_2 = 1/2$, we have the result $\mathcal{P}_1/\mathcal{P}_2 = \mathcal{P} = (1/2)^N$ of Chapter 21, Section 6.

To see how these probabilities compare when we deal with macroscopic quantities of gas, consider for example $V_1 = 0.99V_2$ and $N \cong 10^{23}$. Then $\mathcal{P}_1/\mathcal{P}_2 = 0.99^{10^{23}}$ which is zero for all practical purposes. To see exactly how small this is let us take the logarithm of both sides:

$$\begin{aligned} \log\ (\mathcal{P}_1/\mathcal{P}_2) &= 10^{23}\log 0.99 \approx 10^{23}(-0.01) \\ &\approx -10^{21} \\ \text{or} \qquad \mathcal{P}_1/\mathcal{P}_2 &\approx 10^{-10^{21}} \end{aligned}$$

Thus when a gas expands freely from V_1 to V_2 it will never go back to V_1 by itself. How can we account then for the isothermal reversible oscillation where the volume of the gas increases and decreases periodically at constant temperature? Apparently in the isothermal reversible expansion something must happen which compensates for this enormous change in probability.

In our oscillator we can distinguish two parts, the gas and the heat bath. If the gas does not react chemically with the heat bath, or is not absorbed by it, and if the temperature of the gas and the heat bath are the same, the fact that the gas and the heat bath are in thermal contact does not modify in any way the state of the gas or the state of the heat bath. It follows, therefore, that the probabilities of the states of the gas and of the heat bath can be considered practically independent.



22–12 When a gas expands at constant temperature, the work it does on a piston is equal to the area under the curve P = NkT/V.

This means that if the probability of the macroscopic state of the gas is $\mathcal{P}_{\text{gas}}$ and the probability of the macroscopic state of the heat bath is $\mathcal{P}_{\text{bath}}$, the probability of the macroscopic state of the whole system is

$$\mathcal{P}_{\text{system}} = \mathcal{P}_{\text{gas}} \cdot \mathcal{P}_{\text{bath}}$$

When the gas expands isothermally from a volume V_1 to a volume V_2 the internal energy does not change, but the probability of the macroscopic state of the gas changes from $\mathcal{P}_{1_{\text{gas}}}$ to $\mathcal{P}_{2_{\text{gas}}}$, and a quantity of heat flows from the heat bath through the gas to increase the kinetic energy of the piston and pendulum. This changes the probability of the macroscopic state of the heat bath to $\mathcal{P}_{2_{\text{bath}}}$. But since the process is reversible, the overall probability of the system must remain substantially unchanged:

$$\begin{aligned}\mathcal{P}_{1_{\text{system}}} = \mathcal{P}_{2_{\text{system}}} &= \mathcal{P}_{2_{\text{gas}}} \cdot \mathcal{P}_{2_{\text{bath}}} \\ &= \mathcal{P}_{1_{\text{gas}}} \cdot \mathcal{P}_{1_{\text{bath}}}\end{aligned}$$

Writing it differently,

$$\frac{\mathcal{P}_{2_{\text{gas}}}}{\mathcal{P}_{1_{\text{gas}}}} = \frac{\mathcal{P}_{1_{\text{bath}}}}{\mathcal{P}_{2_{\text{bath}}}}$$

The ratio $\mathcal{P}_{2_{\text{gas}}}/\mathcal{P}_{1_{\text{gas}}}$ for the isothermal reversible expansion is the same as for the free expansion: In both cases the temperature of the gas remains fixed and the internal energy does not change. Hence, for the isothermal reversible expansion,

$$\frac{\mathcal{P}_{1_{\text{bath}}}}{\mathcal{P}_{2_{\text{bath}}}} = \frac{\mathcal{P}_{2_{\text{gas}}}}{\mathcal{P}_{1_{\text{gas}}}} = \left(\frac{V_2}{V_1}\right)^N$$

Thus we have found how the probability of the macroscopic state of the heat bath changes in our reversible process. It has changed because a certain quantity of heat Q flows out from the bath. It does not make any difference whether this heat was absorbed by an expanding gas and transferred to the pendulum or whether it was used up in some other way. In other words, we should be able to express the change in the probability of the heat bath in terms of the quantities which describe the change in the heat bath, that is, Q, and its absolute temperature T. In fact, we can express V_2/V_1 in terms of Q and T. This can be done as follows. Since the temperature of the gas remains constant during the expansion, the heat Q leaving the heat bath is completely converted into mechanical energy of the piston. This increase in mechanical energy equals the work done on the piston by the expanding gas.

The force acting on the piston equals the pressure times the surface area of the piston. If the pressure remained constant during the expansion, the work done would be

$$\begin{aligned}\text{work} &= \text{pressure} \times \text{area} \times \text{distance traveled} \\ &= \text{pressure} \times \text{increase in volume.}\end{aligned}$$

We know, of course, that the pressure, P, decreases according to the relation $P = NkT/V$ where k is Boltzmann's constant. Thus, the work equals the area under the P vs. V curve shown in Fig. 22 – 12. Using integral calculus (see Box) we can show that this area and hence the work is

$$\text{work} = NkT \ln \frac{V_2}{V_1}$$

Here ln stands for natural logarithm. This work is equal to the amount of heat, Q, which was transferred from the heat bath. Thus

$$Q = NkT \ln \frac{V_2}{V_1} = kT \ln \left(\frac{V_2}{V_1}\right)^N$$

We have found earlier that

$$\frac{\mathcal{P}_{1_{\text{bath}}}}{\mathcal{P}_{2_{\text{bath}}}} = \left(\frac{V_2}{V_1}\right)^N$$

Substituting for $(V_2/V_1)^N$ in the above equation for Q gives: $Q = kT \ln \dfrac{\mathcal{P}_{1_{\text{bath}}}}{\mathcal{P}_{2_{\text{bath}}}}$ and we have

$$\ln \frac{\mathcal{P}_{1_{\text{bath}}}}{\mathcal{P}_{2_{\text{bath}}}} = \frac{Q}{kT}$$

The Work Done by a Gas Expanding at Constant Temperature

If, when a gas expands, the increase in volume, ΔV, is small compared to V, the pressure remains practically constant and the work done is

$$\Delta W = P\Delta V$$

The total work done in expanding from V_1 to V_2 is the sum of all the small ΔW's (the area under the curve in Fig. 22 - 12 from V_1 to V_2):

$$W = \sum_{V_1}^{V_2} P\Delta V$$

Since P changes significantly during this large change in volume, we replace P by $P = NkT/V$ and we have

$$W = \sum_{V_1}^{V_2} (NkT/V)\ \Delta V$$

Since N, k, and T do not change,

$$W = NkT \sum_{V_1}^{V_2} \frac{\Delta V}{V}$$

Using integral calculus, we can find the area under the curve exactly. In the notation of calculus

$$W = NkT \int_{V_1}^{V_2} \frac{dV}{V}$$

Integrating between V_1 and V_2 gives

$$W = NkT \ln \frac{V_2}{V_1}$$

This is the relation we have been looking for. It relates the ratio of the probabilities of the initial and final macroscopic states of the heat bath to the quantity of heat which flows out of the bath, and the absolute temperature of the bath.

Notice that the nature of the heat bath does not appear in this relation. It is quite irrelevant whether the heat flows out from a stack of aluminum sheets, a slab of iron, or a container full of water and ice. All that counts is that the body serving as a heat bath has a large enough heat capacity so that it can give off or take in a quantity of heat Q with a negligible change in temperature or none at all, and that the body does not interact with its surroundings.

22–6 ENTROPY

In asking how much more probable one macroscopic state is than another, we ask for a ratio of probabilities. In the last section we have seen that it is more convenient to use the logarithm of this ratio, which equals the difference of the logarithms of the probabilities. For the two cases which we investigated in detail, we found

$$\ln \frac{\mathcal{P}_{2_{\text{gas}}}}{\mathcal{P}_{1_{\text{gas}}}} = \ln \mathcal{P}_{2_{\text{gas}}} - \ln \mathcal{P}_{1_{\text{gas}}} = N \ln \frac{V_2}{V_1}$$

for the free expansion of a gas at constant temperature, and

$$\ln \frac{\mathcal{P}_{1\,\text{bath}}}{\mathcal{P}_{2\,\text{bath}}} = \ln \mathcal{P}_{1_{\text{bath}}} - \ln \mathcal{P}_{2_{\text{bath}}} = \frac{Q}{kT}$$

for the loss of a quantity of heat Q by a heat bath at absolute temperature T. We shall give the natural logarithm of the probability of a macroscopic state of a system the name *entropy** and designate it by the letter S:

$$S = \ln \mathcal{P}$$

Then the first two equations in this section can be written as

$$\Delta S_{\text{gas}} = N \ln \frac{V_2}{V_1}$$

and

$$\Delta S_{\text{bath}} = -\frac{Q}{kT}$$

Notice that the measurable changes in a system, such as the change in volume or the amount of heat transferred to it, give us directly the change in entropy. Thus we cannot calculate the entropies of the initial and final states themselves.

---

*In most textbooks the entropy is defined as $k \ln \mathcal{P}$, but we prefer to keep it a pure number.

For two systems A and B which are independent, the total probability is $\mathcal{P} = \mathcal{P}_A \mathcal{P}_B$. This means that for the corresponding entropies we have

$$S = S_A + S_B$$

and for the changes in entropy

$$\Delta S = \Delta S_A + \Delta S_B$$

For a reversible process the total probability remains unchanged; hence

$$\Delta S = 0$$

and therefore

$$\Delta S_B = - \Delta S_A$$

Applying this to the isothermal reversible oscillation of our gas oscillator, we say that the increase of the entropy of the gas during the expansion part equals the decrease of entropy of the heat bath while it is losing heat to the gas.

22–7 A REALISTIC HEAT BATH

In the gas oscillator we were able to reduce the temperature changes by increasing the number of aluminum sheets. To keep the temperature change really down to zero, we would have to use an infinite number of aluminum sheets—which is, of course, impossible. Nevertheless, we can construct a realistic heat bath which will stay at constant temperature while absorbing or giving off finite quantities of heat. A mixture of water and ice acts as a heat bath at 273°K as long as we do not freeze all the water or melt all the ice. If we had used a water-ice mixture as the heat bath for our gas oscillator, the temperature would have remained at 273°K, provided the container was a good enough conductor so that the temperature remained constant while the piston moved up and down.

What happens as the piston pumps heat in and out of our bath reversibly? Some of the ice melts on the downstroke and then freezes again on the upstroke. Because the temperature is truly constant the entropy of the bath first increases by Q/kT and then decreases by the same amount.

For this system it is clear what the change in entropy is associated with. When the entropy is increasing, ice is melting to water; when it is decreasing, water is freezing into ice. We lose entropy in making ice. This is consistent with our interpretation of the entropy as being the logarithm of the probability. Clearly, the beautiful, ordered crystalline structure of ice is much less probable than the more random distribution of molecules in water, even at the same temperature.

22–8 ENTROPY CHANGES IN AN INELASTIC COLLISION

We have seen that when a body with a large heat capacity absorbs a quantity of heat Q at absolute temperature T its increase in entropy is

$$\Delta S = Q/kT$$

Since the change in entropy depends only on the initial and final state of the system, it is irrelevant how this quantity of heat was transferred to the body; it could be absorbed through contact with another body at a higher temperature, or the heat could be produced by dissipating mechanical energy.

Consider two putty balls each of mass m and speed v colliding head on and sticking together. In this case, the increase in internal energy is the same as if a quantity of heat $Q = mv^2$ had been added to the bodies. If the heat capacity of the balls is large enough, the increase in temperature is insignificant. How likely is it that after the collision a fluctuation in the distribution of the momenta of the atoms will occur which will convert the increase in internal energy back into the original kinetic energy of the two balls so that they fly apart with their initial speeds?

For this process

$$\Delta S = -\frac{Q}{kT} = -\frac{mv^2}{kT}$$

To get an idea of the order of the magnitude of entropy changes let us take $m = 1$ kg and $v = 1$ m/sec. Hence $mv^2 = 1$ joule. Taking $T = 300°\text{K}$ and recalling that $k \approx 2 \times 10^{-23}$ joule/°K we find

$$\Delta S \approx -10^{20}$$

This is a large decrease in entropy. But the change in entropy is the logarithm of the ratio of the two probabilities of the macroscopic states of the system. This ratio is thus $e^{-10^{20}} \approx 10^{-10^{20}}$, which again is a long way of saying "zero."

Notice that while in the expression for the change in the entropy in the expansion of a gas, $\Delta S = N \ln(V_2/V_1)$, the number of atoms appeared explicitly, it seems to be absent in the expression for the change of entropy in heat transfer. Actually the number of atoms comes in very distinctly. If we express the mass of the moving balls as the product of the number of atoms (N) times the mass of one atom (m_a), then

$$\Delta S = -N \frac{m_a v^2}{kT}$$

For room temperature and atoms of medium mass,

$$\frac{m_a v^2}{kT} \approx 10^{-4}$$

For $N \approx 10^{24}$ this gives $\Delta S = -10^{20}$, the result we had before.

22–9 ENTROPY CHANGES IN HEAT CONDUCTION AT CONSTANT VOLUME

Consider a large insulated container which is divided into two equal compartments by a glass partition. If we fill the two compartments with equal masses of water at temperature T_1 and T_2 where $T_1 < T_2$, and stir the water continually, then the temperature in each part of the container will remain uniform while heat leaks across the wall. Let us calculate the change in entropy as a small quantity of heat Q is transferred from the warmer compartment to the colder with only a negligible change in temperature. The cooler water gains a quantity of heat Q at a temperature T_1; thus its entropy changes by Q/kT_1. Meanwhile the warm water loses a quantity of heat Q at a temperature T_2, so its entropy changes by $-Q/kT_2$. The total change is therefore

$$\Delta S = \frac{Q}{kT_1} - \frac{Q}{kT_2} = \frac{Q(T_2 - T_1)}{kT_1T_2}$$

Since $T_2 > T_1$, ΔS is always positive. Recalling that an increase in entropy means an increase in the probability of the macroscopic state of the system, we see that the flow of heat from a warm to a cold body is indeed irreversible. Heat does not pass spontaneously from a colder body to a warmer body.

We can also find the change in entropy when the heat transferred to a body is large enough to cause a considerable change in temperature. The method is similar to that which we used in calculating the work done by a gas pushing a piston at constant temperature from a volume V_1 to a volume V_2. Consider a body at constant volume whose heat capacity, C, is independent of temperature. As heat flows into it, equal quantities of heat ΔQ raise its temperature by equal amounts $\Delta T = \Delta Q/C$. Each small change in temperature changes the entropy by

$$\frac{\Delta Q}{kT} = \frac{C\Delta T}{kT}$$

The total change in entropy as the body's temperature changes from T_1 to T_2 is then given by the sum

$$\Delta S = \sum_{T_1}^{T_2} \frac{C\Delta T}{kT} = \frac{C}{k} \sum_{T_1}^{T_2} \frac{\Delta T}{T}$$

This is the same kind of sum we calculated in the Box and yields

$$\Delta S = \frac{C}{k} \ln \frac{T_2}{T_1}$$

We can apply the result to find the change in entropy when two equal masses of water in the container described at the beginning of this section are allowed to remain in thermal contact long enough for their temperatures to become equal. Let the initial temperatures of the two masses be $T + \Delta T$ and $T - \Delta T$ and their final temperature T. The total change in entropy is then

$$\Delta S = \frac{C}{k}\left[\ln \frac{T}{T - \Delta T} + \ln \frac{T}{T + \Delta T}\right]$$

Since $\frac{T}{T - \Delta T} > 1$, the first term is positive and represents the increase of entropy of the initially colder mass. The ratio $\frac{T}{T + \Delta T} < 1$ and so its logarithm is negative. This represents the decrease in entropy of the initially warmer mass.

Rewriting the expression for ΔS, we have

$$\Delta S = \frac{C}{k} \ln \frac{T^2}{(T-\Delta T)(T+\Delta T)} = \frac{C}{k} \ln \frac{T^2}{T^2-(\Delta T)^2}$$

Since $T^2 > T^2 - \Delta T^2$, the logarithm is positive and the entropy increases, showing again that the macroscopic state of equal temperature is the most probable one of the system.

An example of a system with a constant heat capacity is an ideal monatomic gas at constant volume. Since the internal energy is $E = (3/2)NkT$, the heat capacity, which is the change in internal energy per degree change in temperature, is $C = (3/2)Nk$. The change of entropy of an ideal gas when its temperature changes from T_1 to T_2 is then

$$\Delta S = \frac{3}{2} N \ln \frac{T_2}{T_1}$$

The total change in entropy when one mass of gas cools from $T + \Delta T$ to T and is in thermal contact with another equal mass of gas which warms from $T - \Delta T$ to T is

$$\Delta S = \frac{3}{2} N \ln \left(\frac{T^2}{T^2-(\Delta T)^2}\right)$$

Here we see once again that the number of atoms appears explicitly in the change of entropy. For any macroscopic quantity of gas and the smallest measurable difference in temperature, ΔS will be very large. (See Question 23.)

For any substance—solid, liquid, or gas—the heat capacity is, of course, proportional to the number of atoms. In general the heat capacity per atom is not constant, but it is never very different from k. Thus the relation

$$\Delta S \approx N \ln \frac{T^2}{T^2-(\Delta T)^2}$$

for the changes in entropy when two equal masses are in thermal contact and come to the same temperature holds also for liquids or solids.

22–10 THE ENTROPY OF AN IDEAL GAS WHEN BOTH VOLUME AND TEMPERATURE CHANGE

Consider a sample of N atoms of an ideal monatomic gas which at one time has a volume V_1 and absolute temperature T_1 and at a later time has a volume V_2 and temperature T_2. What is the total change in entropy? If at both times the gas is in equilibrium, that is, the temperature and pressure are uniform throughout and there is no bulk motion of the gas, then the probability of an atom having a certain kinetic energy is independent of its position. Furthermore, since in an ideal gas there are no forces between the atoms, the probability of an atom having a given kinetic energy is the same whether other atoms are nearby or not. Therefore, $\mathcal{P}_T$, the part of the probability of the macroscopic state of the gas which depends on the temperature, is independent of $\mathcal{P}_V$, the part which depends on the volume. In other words, we can treat the two parts of the probabilities as belonging to two independent systems and the total probability is

$$\mathcal{P} = \mathcal{P}_V \mathcal{P}_T$$

For the total entropy of the gas it then follows (see Section 22–6) that

$$S = S_V + S_T$$

and for the change in entropy

$$\Delta S = \Delta S_V + \Delta S_T$$

We have already found separate expressions for ΔS_V and ΔS_T. They are

$$\Delta S_V = N \ln \frac{V_2}{V_1} \text{ and } \Delta S_T = \frac{3}{2} N \ln \frac{T_2}{T_1}$$

Thus,

$$\Delta S = N \ln \frac{V_2 T_2^{3/2}}{V_1 T_1^{3/2}}$$

This is the general expression for the change of entropy in an ideal monatomic gas when both volume and temperature change.

For an adiabatic reversible expansion $\Delta S = 0$; hence for a monatomic gas

$$V_2 T_2^{3/2} = V_1 T_1^{3/2}$$

or, in other words,

$$VT^{3/2} = \text{constant}$$

Using the ideal gas law $PV = NkT$, we can also see that

$$V\left(\frac{PV}{Nk}\right)^{3/2} = \text{constant}$$

or

$$PV^{5/3} = \text{constant}$$

This last equation is the standard way of expressing the relationship among the parameters for the adiabatic expansion of an ideal monatomic gas and is called the adiabatic gas law. Of course, it can be derived without using the concept of entropy. After all, we know the change in internal energy of the gas in going from one temperature to another. This must be equal to the work done on the piston, and this can be used to derive the relations above.

22–11 DIFFUSION—THE ENTROPY OF A PERFECT GAS

In Section 22 – 5 we studied the free expansion of an ideal gas. Let us apply what we learned in that section to the free expansion of one gas into another—the diffusion of a gas into another. We can calculate the entropy change in this process, and in so doing we shall learn more about the entropy of ideal gases.

Imagine a closed container of volume V containing a partition that divides it into two equal volumes $V/2$. Suppose one half of the container is filled with N atoms of argon gas and the other half contains N atoms of neon gas. Since the mass of an argon atom is different from the mass of a neon atom, the density of the gas in the two halves is different.

What happens when we take away the partition between the gases? Quite naturally they begin to mix, and after a while they are both distributed uniformly over the total volume. Each gas has diffused into the other, and the density is now the same in both halves. It is also quite clear that the two different gases will never separate again into two halves, one pure argon and the other pure neon. In other words, there has been a large increase in the entropy during the diffusion process.

To calculate the change in entropy we need only realize that what has occurred is a free expansion of each of the gases separately into twice the volume either of them occupied originally. Since the atoms of neon and argon do not interact with each other, the total change in entropy is the sum of the changes in entropy of the two gases. Since in this case $V_2 = 2V_1$, then

$$\Delta S = \Delta S_{\text{argon}} + \Delta S_{\text{neon}}$$

$$= N \ln \frac{2V_1}{V_1} + N \ln \frac{2V_1}{V_1} = 2N \ln 2$$

which is an enormous change in entropy for any macroscopic samples of gas.

Suppose, now, that both compartments initially contained neon. What happens to the entropy when we remove the partition? As before, the atoms from each half will spread through the total volume. But we cannot tell the atoms from the two halves apart since they are all atoms of the same kind of gas, and so we can observe no change after the partition is removed. The density remains the same, and if the gas had a color it would not change. In short, the gas is in the same macroscopic state as it was before the partition was removed, and since there is no change in the macroscopic state of the gas there can be no change in its entropy. The final entropy, the entropy of the whole volume after the partition is removed, must be the same as the sum of the entropies of the two halves before the partition is removed. But if we calculate the change in entropy in the same way as we did when two different gases diffused into each other, we get the same answer we did before: $\Delta S = 2N \ln 2$ and we know this cannot be right. Let us try to find a form for the entropy function that will give what we know is the correct answer for this case as well as for the mixing of two different gases.

If we consider the average volume occupied per atom of each gas in the two processes we have just analyzed, we shall see a possible solution to this apparent paradox. In the process in which equal volumes of neon and argon were allowed to mix, the average volume per atom in each compartment before the mixing was $\left(\frac{V_1}{N}\right)_{\text{neon}} = \left(\frac{V_1}{N}\right)_{\text{argon}}$, but after mixing, each gas occupied a volume $V_2 = 2V_1$ and the average volume occupied by one atom of each gas was larger and equal to $\frac{2V_1}{N}$. In this case, we found an increase in entropy.

On the other hand, in the case of equal volumes of the same kind of gas, the average volume per atom of the N atoms in each compartment before the partition was removed was also V_1/N, but after the partition was removed, since only one kind of gas was present, it was $V_2/2N = 2V_1/2N = V_1/N$, the same as before the partition was removed. In this case the average volume per atom of the gas does not change and, as we know, there is no change in entropy. This suggests that entropy and probability should be expressed in terms of the average volume occupied by a single atom. We shall now rewrite our expressions for the entropy of a freely expanding gas in terms of the average volume per atom, V/N.

In Section 22–5 we found the probability of the macroscopic state of a gas occupying a volume V was $(V/V_0)^N$ where V_0 was a large reference volume. Replacing V in this expression by the V/N gives for the probability

$$\mathcal{P} = \left(\frac{V}{NV_0}\right)^N$$

and for the entropy we have

$$S = \ln \mathcal{P} = N \ln \left(\frac{V}{NV_0}\right)$$

Applying this function to the mixing of argon and neon we have

$$\Delta S = \Delta S_{\text{argon}} + \Delta S_{\text{neon}}$$

Since the change in entropy of the argon is equal to that of the neon, the total change is

$$\begin{aligned}\Delta S &= 2(N \ln 2V/NV_0 - N \ln V_1/NV_0)\\ &= 2\,N \ln 2\end{aligned}$$

the same result we obtained before.

For the case where the same kind of gas was in each compartment we have

$$\begin{aligned}\Delta S &= S_2 - S_1\\ &= (2N) \ln 2V_1/(2N)\,V_0 - 2(N \ln V_1/NV_0) = 0\end{aligned}$$

as we know it must be.

One more remark about our new expression for the volume-dependent part of the entropy of an ideal gas is in order. We can rewrite the equation in the following form

$$S = N \ln \left(\frac{V}{V_0}\right) - N \ln N$$

which gives the correct form for the entropy equation of a gas when its volume alone changes. This equation contains a constant, independent of the volume changes but dependent on N, the number of atoms in the system.

It is worth noticing why we have not needed this constant until now. The reason is that in all previous processes the number of atoms of one kind in each system remained unchanged, and since we always calculated changes in entropy, the constant $N \ln N$ dropped out. In the last case we started with two systems, each containing N atoms of the same kind, and ended up with a system containing $2N$ atoms of the same kind.

22–12 THE SECOND LAW OF THERMODYNAMICS

At the beginning of the preceding chapter we described a variety of irreversible processes. We have seen in this chapter that all these processes have one thing in common: in all of them the entropy of the system increases. In special limiting cases we found very nearly reversible processes, and the entropy of the system remained substantially the same. The generalization of these observations is known as the second law of thermodynamics: In processes occurring spontaneously the entropy of the system never decreases except for small fluctuations.[*]

In applying this law we have to be careful to include in the system any part of the universe which is affected by the process under consideration. For example, consider the freezing of a mass of water. The entropy of the water decreases as it gives off its heat of fusion to the surroundings. So we have to include the surroundings in the system, and we already know that in this case the total entropy increases since heat flows from the water at 0°C to the colder surroundings.

In many cases where the entropy of part of a system decreases, it is not so easy to see that the overall entropy increases, as in the case of the freezing of water on a cold day. Consider an electric refrigerator, for example. When we turn it on, the lower temperature inside decreases while the higher temperature outside increases; thus, the entropy of refrigerator and room decreases! But the refrigerator, of course, does not work by itself. It has an electric motor which must be connected to a power line which in turn is connected to a power plant in which some fuel is burned in a chemical reaction to produce steam which makes the turbines turn. To follow the overall change in entropy we have to learn how to find the change in entropy when a chemical reaction takes place. When we do this, we find that the entropy of the whole system increases.

---

[*] The first law of thermodynamics is the extension of the law of conservation of mechanical energy to include internal thermal energy (See Chapter 20, Section 8.)

There is another extension of the idea of entropy which we need if we want to account for processes where heat is transferred by radiation rather than by conduction. You may recall that the conservation of momentum seemed to be violated when we considered only the momentum of material particles in processes where light was emitted or absorbed. (See Chapter 16, Section 7.) When we included the momentum of light, the validity of the law was restored. We had similar results with the conservation of angular momentum (Chapter 19, Section 11) and the conservation of energy. The situation is quite similar with respect to entropy. Stars give off huge amounts of energy. Eventually they must cool off, and so their entropy decreases. Is this in violation of the second law of thermodynamics?

As the star cools, more and more radiation spreads throughout space. If the increase of entropy of the radiation is taken into account, it is found that the total entropy increases in agreement with the second law.

The growth of living organisms on our earth seems to constitute a violation of the second law. The entropy of the water and minerals in the soil and the carbon dioxide decreases when they form part of a plant. The plant presents a much more ordered and therefore a less probable macroscopic state than the state in which the ingredients were before. On the one hand, this process can take place only with the help of radiation from the hot sun. On the other hand, the earth can absorb heat from the sun without its temperature rising only because it radiates heat into the cold outer space. If we take into account the entropy of the radiation absorbed and emitted by the earth as well as the entropy of the growing plant and its surroundings, it is found that the total entropy increases.

QUESTIONS

1.* How do you describe the macroscopic state of the marble machine discussed in the preceding chapter?

2.* Five hundred people are attending a large banquet. Which of the following describe the "macroscopic state" of the banquet hall and which describe the "microscopic state"?

(a) The number of men and the number of women.

(b) The names of the people at the head table.

(c) The amount of champagne consumed by each person.

(d) The average amount of champagne consumed by each person.

3.* The damping time of a certain pendulum in air is T. Would the damping time in a vacuum be greater or less than T?

4. The motion of a damped pendulum is irreversible. Suggest several examples of irreversible and almost reversible processes from everyday life.

5.* Why is there a temperature difference between the gas and the container in the gas oscillator?

6.* What limiting condition on the surface area of the gas oscillator will give the maximum temperature change of the gas between maximum and minimum volume?

7.* Suppose the walls of the cylinder in the gas oscillator were made of a nearly perfect heat insulator. What would be the damping coefficient of the system if this were the case and no aluminum sheets were in the cylinder?

8. Aluminum is a good conductor of heat. What would be the effect on the damping of the gas oscillator described in Section 22–3 if plastic foil (a poor heat conductor) of the same heat capacity as the aluminum had been used?

9. The cylinder of the gas oscillator described in Section 22–3 contained about 29 gm of air. The aluminum sheets in the cylinder each had a mass of about 4.8 gm.

   (a) How much energy would be necessary to change the temperature of the air in the cylinder by 1°C? (About 1 joule is needed to raise the temperature of 1 gm of air 1°C in this case.)

   (b) What temperature change would occur in 400 of the aluminum sheets if they absorbed the amount of energy needed to raise the temperature of the air in the cylinder 1°C? (About 1 joule is needed to raise the temperature of 1 gm of aluminum 1°C.)

10. If the piston in the cylinder in Fig. 22–3 is given a push, it will oscillate up and down. Assuming that no heat can leak through the walls and that none is absorbed by the walls, describe the energy changes in the system through one complete cycle:

    (a) when there are no aluminum sheets in the cylinder.

    (b) when there are enough aluminum sheets to cause considerable damping.

11. Does the position in which the piston comes to rest in part (b) of the preceding problem coincide with the middle point of the oscillation in part (a)?

12.* (a) If the N molecules of a gas must be somewhere in volume V_0, which contains volume V, what is the probability of finding no molecules at all in V?

    (b) Under what conditions will this probability be greater than the probability of finding *all* the molecules in V?

13.* A system is at equilibrium. What is the probability that its entropy will spontaneously *decrease* by ΔS?

14. A common type of radio vacuum tube has a volume of about 10 cm^3 and the gas pressure inside is only about 10^{-8} atmosphere. What would be the decrease in entropy of the gas if, owing to a fluctuation, it should occupy only 0.999 of the volume? How many times would you have to examine the tube in order to have a reasonable chance of observing such a fluctuation?

15.* In general, which has a higher entropy, a given mass of liquid or the vapor of the same substance at the same temperature?

16.* What is the change in entropy of an ice-water mixture when 20 gm of ice melts? (It requires about 335 joules to melt 1 gm of ice.)

17. In Section 22–8, we found that the probability that two 1-kg putty balls in contact at room temperature will spontaneously fly apart with speeds of 1 m/sec is about $10^{-10^{20}}$. If you were to try to write this number without using powers of ten, it would fill many, many books. In fact, the Library of Congress, which holds about 10^7 books, could store only 10^{13} digits. Express in powers of ten the number of libraries the size of the Library of Congress that would be needed to store $10^{-10^{20}}$ when written without using powers of ten.

18.* A simple pendulum of mass m and length l is initially displaced by a small angle θ. What is the increase in entropy of the surrounding air when the oscillation has damped out?

19. When a small snowball strikes the ground, the impact will turn its kinetic energy into internal energy. If the speed were high enough, this would completely melt the snowball. What is the change in entropy in the process? What assumptions have you made in your calculation? What is the probability that this puddle of water will suddenly freeze and leave the ground with the speed at which the snowball struck?

20. If, as a result of fluctuations of the distribution of the momenta of the atoms in your body, you should rise 1 cm vertically out of your chair, what drop in your body temperature would occur? Estimate how likely this event is.

21. Suppose 2 kg of water at 300.3°K is in thermal contact with 1 kg of water at 300.0°K. (a) What change in entropy will occur? (b) What is the probability that, after thermal equilibrium has been reached, a large fluctuation will take place so that the 2 kg of water will warm back to 300.3° K and the 1 kg of water cool back to 300.0°K? The specific heat of water is about 4 joules/(gm°C).

22. Suppose you wish to cool a piece of iron from temperature T to $T - \Delta T$. Will you have to "pay less" in entropy if you bring the iron in contact (a) with a heat bath at temperature $T - \Delta T$, or (b) with another piece of iron of equal mass at temperature $T - 2\,\Delta T$?

23. Consider an ideal monatomic gas of $2N$ particles in a volume $2V$ at temperature T.

    (a) What is the change in entropy if one half of the gas increases its temperature by ΔT, while the other half decreases by ΔT?

(b) Calculate the change in entropy for the special case where $(\Delta T/T) \ll 1$ and use the approximation $\ln(1 + x) \simeq x$ for $x \ll 1$.

(c) If $N = 10^{24}$, how small must $\Delta T/T$ be so that the probability is not much less than 1?

(d) If $(\Delta T/T)$ is to be 10^{-3}, what must be the value of N to have a probability not much less than 1? To what volume does this correspond at room temperature and atmospheric pressure?

24. Consider an insulated container which has two compartments separated by a thin wall. In one compartment there is a mixture of solid and liquid naphthalene at 80°C (the melting point of naphthalene). In the other there is a mixture of ice and water at 0°C. Describe what will happen inside the container as time goes on.

25. For which reversible processes does

$$PV^{5/3} = \text{constant}$$

apply to an ideal monatomic gas and when does

$$PV = \text{constant}$$

apply?

26. In an adiabatic expansion of an ideal monatomic gas, the relation between pressure and temperature is independent of the volume. What is the form of the relation?

27. When a west wind blows over a mountain range that runs north and south, air from the valley floor on the west flows over the mountain and down to the same level on the other side. As the air rises it expands, owing to the lower atmospheric pressure at higher altitudes, but no heat is exchanged with the surrounding atmosphere.

(a) How does the temperature of the air change as it flows up over the mountain and down the other side?

(b) Is the process reversible?

28. Suppose the air described in Question 27 contains enough water vapor so that condensation in the form of rain begins as it approaches the summit from the west side.

(a) Describe qualitatively the changes in entropy and temperature of the air as it flows from one valley floor to the other.

(b) Is the process reversible?

29. Imagine a gas of $2N$ particles in a volume $2V$. Normally the number of particles in each half of the volume would be N.

(a) Calculate the probability for the gas to be in this state using the general expression

$$\mathcal{P} = \left(\frac{V}{NV_0}\right)^N$$

good for any volume and number.

(b) Suppose that one half of the volume contains $(N + n)$ particles while the other contains $(N - n)$ particles. Compare the probability of this with your answer in (a).

(c) From the ratio of the probabilities calculate the change in entropy in going from the (N, N) state to the $(N + n)$, $(N - n)$ state for $n/N \ll 1$ using the approximation

$$\ln(1 + x) \simeq x \text{ when } x \ll 1.$$

(d) If $N = 10^{24}$, how large can n become without the ratio of the probabilities getting much less than 1? What would be the ratio of the densities of the parts of the gas for such values of n?

30. Consider a volume of gas V broken up into n compartments. What is the entropy of the gas expressed as the sum of the entropies of the n compartments?

31. Two containers separated by a partition both contain air at room temperature and atmospheric pressure. The volume of one container is V and the volume of the other is $2V$.

(a) What will be the change in entropy of the air if the partition is removed?

(b) What entropy change would occur if each container held a different gas?

32. Suppose we have three different gases in separate compartments of equal volume, at the same temperature and pressure. What is the change in entropy if we let the three gases mix by taking out the partitions?

23 Some Qualitative Facts About Electricity

WE HAVE learned that every material body attracts every other material body with a force known as the *gravitational force*. This gravitational attraction has practical consequences only when at least one of the interacting bodies has enormous mass, something like a whole planet. However, gravitational forces are not the only forces acting at a distance between material bodies. Sometimes other forces are enormously greater. A small magnet will lift a steel nail off the table against the gravitational attraction of the entire earth. A comb rubbed on your sleeve will lift bits of paper. These are examples of *magnetic* and *electric* forces, respectively.

Although the existence of such forces has been known since antiquity, it was only during the Renaissance that a systematic study of electricity and magnetism developed, and physicists did not gain a clear understanding of these subjects until the end of the past century. Hardly ever has a scientific achievement had such profound and far-reaching consequences. There have been innumerable practical applications. The harnessing of electric power and the development of electrical communications have changed our whole way of life. On the scientific side, we have learned that electric forces control the structure of atoms and molecules. Electricity is associated with many biological processes—for instance, with the action of our nerves and brain.

23–1 ATTRACTION AND REPULSION BETWEEN ELECTRIFIED OBJECTS

We shall now examine some of the basic facts of electric and magnetic behavior and discuss their interpretation. Let us begin with a simple electrical experiment. Rub a glass rod with a piece of silk and then place it horizontally in a stirrup hung on a silk thread. Then rub a second glass rod, and bring it near the first. The two rods *repel* each other.

Repeat this experiment with two rods of plastic rubbed with a piece of fur. These two rods repel as the two glass rods did (Fig. 23–1).

Finally rub a glass rod with silk and a plastic rod with fur. Place one rod in the stirrup and bring the other near it. We now find that these rods *attract* each other.

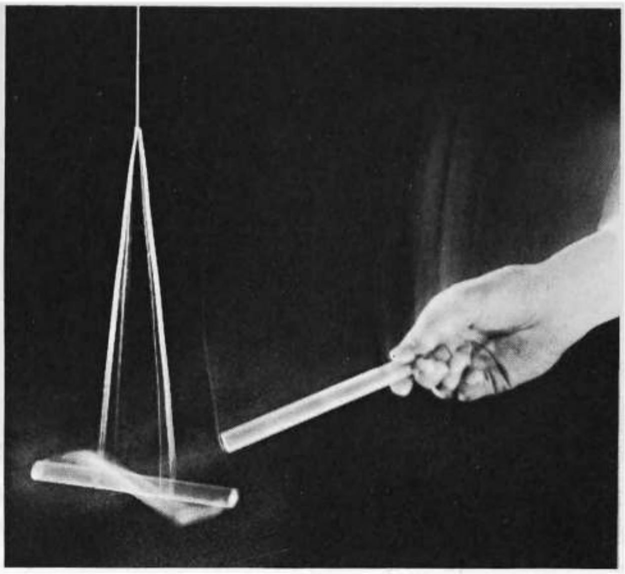
We can perform similar experiments with a number of other materials. Objects of the same material which have been "electrified" by the same procedure always repel. Different electrified objects may either attract or repel.

We find that electrified objects fall into two groups. Only two "electric states" exist, one similar to that of the glass rod and one similar to that of the plastic rod of the previous example. Following the common usage, which dates back to Benjamin Franklin, we shall say that the glass rod and all other electrified objects which behave like it are *positively charged*. Similarly we shall say that the plastic rod and all other electrified objects which behave like it are *negatively charged*.* Any two positively charged objects repel each other, just as two electrified glass rods do. Similarly any two negatively charged objects repel each other. Any positively charged object attracts any negatively charged one.

23–2 ELECTRIC FORCES BETWEEN THE BUILDING BLOCKS OF MATTER

We have learned that matter is made of atoms and that the atoms themselves may be made of smaller units. But so far in this book we have not faced the question of whether there are any ultimate building blocks from which all atoms are made. The study of electricity will give us some important information about this question.

To begin with, it is natural to think that the force between two bodies is the vector sum of the forces acting between their building blocks. Since two electrified objects attract or repel each other, we suppose that there are similar forces of attraction or repulsion between some of the parts of which the two objects are made. Because there are electrically positive and electrically negative objects, we assume that among the building blocks of matter there are two types of particles: *electrically positive* and *electrically negative*. Two particles of the same sign repel and two particles of opposite sign attract. Some of the building blocks of matter belong in a third group: for example, neutrons (Section 17–8) do not act like either positive or negative electric particles.



23–1 Two electrified plastic rods repel each other when they are brought close together. This photograph is a time exposure, made while one rod was brought near the other. The repulsive force pushed away the near end of the suspended rod.

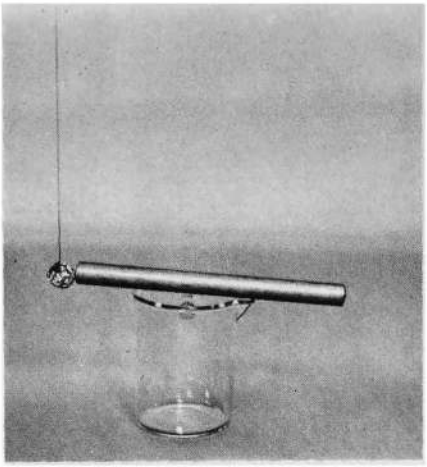
Although the electrical forces between particles are enormously greater than the gravitational forces, when ordinary "uncharged" pieces of matter are placed near each other, they exert no appreciable electric forces on each other. This does not mean we have to give up the idea that there are positive and negative electric particles in matter. Since a positive and a negative particle exert forces in opposite directions on any third electric particle, we merely note that the forces exerted by the two kinds of particle can cancel. When the effects of the positive and negative particles cancel exactly, we say the object is uncharged or electrically neutral. We can even picture a neutral atom in this way.*

If we add some positive particles to an electrically neutral object, there is no longer a balance. The effect of the positive particles is

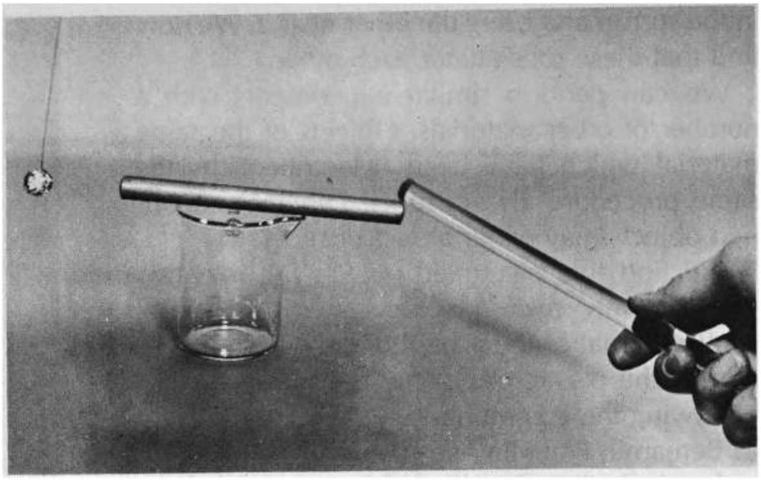
---

*The words "positive" and "negative" are used because the forces exerted by neighboring positively and negatively charged objects on any third body tend to cancel.

*If every small volume of a body is neutral, the resultant force exerted by all the electric particles upon any electrical particle outside is zero. However, a body can be neutral on the average and nevertheless contain local concentrations of charge. Then an electric particle near one of the local concentrations is subject to a force.



23–2 **(a) A metal rod on a beaker is in contact with a light metal-coated ball.**



(b) We touch the end of the metal rod with a charged rod, and the ball at the other end is repelled.

greater than the effect of the negative particles; and we say the object is positively charged. We can also charge an object positively by *removing* some *negative* particles, leaving an excess of positive ones. (Similarly, we can make a neutral object electrically negative either by adding some negative particles to it, or by taking some positive particles away from it.)

We have seen that rubbing a glass rod with silk makes the rod electrically positive. How can this happen? We may think of two possibilities. Perhaps there is a transfer of positive particles from the silk to the rod; perhaps there is a transfer of negative particles from the rod to the silk. In either case the silk must become electrically negative as the glass becomes electrically positive. We can check this conclusion by bringing the silk near the suspended glass rod: We find that the rod is attracted by the silk. Also the silk will repel a negatively charged plastic rod.

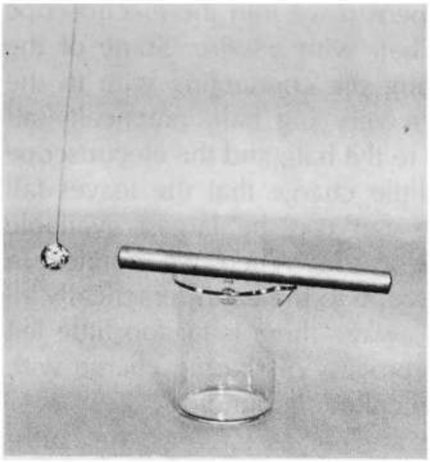
23–3 INSULATORS AND CONDUCTORS

We often classify different materials by saying that one is an electrical conductor and another an insulator. This classification is based on experiments like those we shall now discuss. Suppose we hang a light metal-coated ball on a silk thread. Then we support a metal rod horizontally on a glass or plastic stand so that one end touches the ball. Now we electrify a piece of glass, and rub the glass against the other end of the metal rod. As we rub the glass against the metal, the ball swings away. We repeat the experiment, using a rod of plastic in place of the metal rod. Now the ball does not move (Fig. 23 – 2). We therefore see that the metal and the plastic rods act differently.

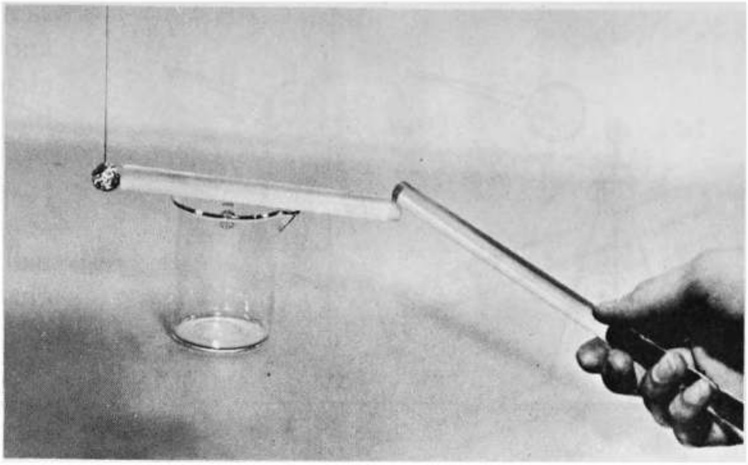
To account for this difference we need only assume that in a metal some electric particles are free to move from point to point, while in the plastic no electric particles can move freely. Suppose, for example, that the free particles in the metal are negative. When the positively charged glass touches the neutral metal rod, the glass pulls some of the free negative particles off the metal rod and the ball. Both the ball and the rod therefore become positively charged, and they repel each other. Even after the glass is removed, there will be a shortage of negative particles on both the rod and the ball, and they will continue to repel.

We can equally well assume that the free particles are electrically positive. Then, when we touch the metal rod with the electrified glass, positive particles will pass from the glass to the metal and spread along it to the ball. Again, the rod and the ball will both become positively charged and will repel each other.

Now let's picture what happens when we replace the metal rod with the plastic one. In the plastic neither positive nor negative particles can move freely. Thus only the part of the rod in direct contact with the charged glass acquires



(c) After the charged rod is removed, the metal rod still repels the ball.



(d) We repeat the experiment with a plastic rod in place of the metal one. The ball stays still.

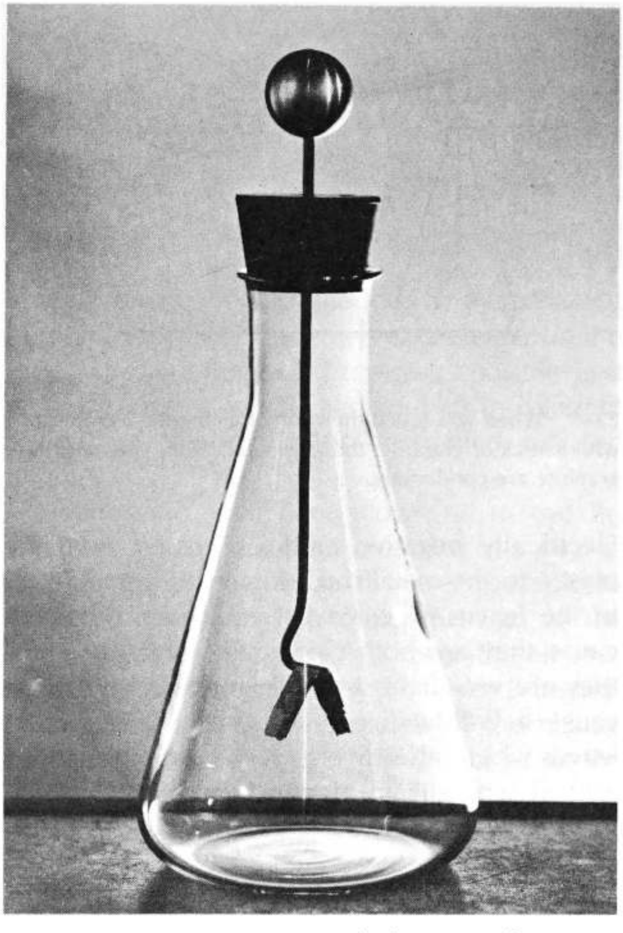
an excess of positive particles. The rest of the rod remains electrically neutral, and so does the ball. Consequently, there is no force pushing the ball away from the rod.

Substances that behave like the metal in this kind of experiment we call *conductors*. Substances that behave like the plastic we call *insulators*. All conductors contain freely moving electric particles and insulators do not.

Toward the end of this chapter we shall learn how to determine the sign of the freely moving particles in different kinds of conductors. We shall find that when liquids and gases conduct electrically, both positive and negative particles move. In metals, however, the conductivity is due solely to the motion of electrically negative particles. To keep the discussion brief, in the next few sections we shall assume that only negative particles move.

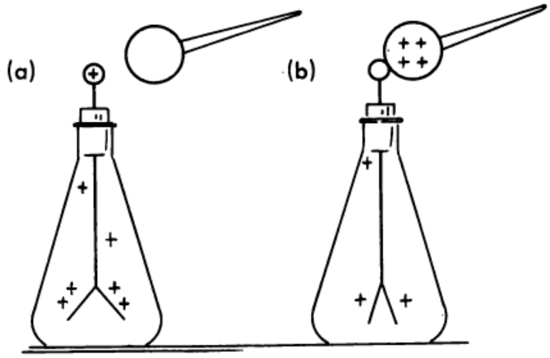
23–4 SOME EXPERIMENTS WITH AN ELECTROSCOPE

An electroscope is an instrument for detecting the presence of electric charge. The simplest version of this instrument is an old-fashioned *gold-leaf electroscope*. It consists of two thin gold leaves hanging from a metal rod in a glass jar (Fig. 23 – 3). We shall now explain how it works.

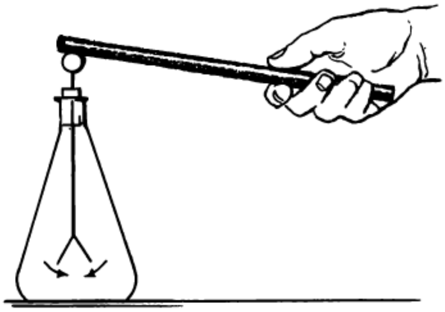


23–3 A gold-leaf electroscope. The leaves stand out because they are similarly charged.

If we touch the knob at the upper end of the metal rod with a charged piece of plastic, the gold leaves spring apart. What has happened?



23–4 (a) A charged electroscope, with the leaves repelling each other strongly. (b) We have touched the knob with a metal ball on an insulating handle. The charge is now shared between the ball and the electroscope; and the leaves have now fallen considerably.



23–5 When you touch the knob of a charged electroscope with a stick of graphite, the leaves fall. Both you and the graphite are conductors.

Electrically negative particles passed from the plastic to the metal rod and spread immediately to the leaves. The leaves repel each other because they are both electrically negative. Since they are very light, a small charge is enough to cause a visible divergence of the leaves. (The leaves would also diverge if we used a glass rod rubbed with silk in place of the plastic rod. In this case, however, the leaves would be electrically positive.)

Give a charge to an electroscope and then touch the knob with an uncharged metal ball on an insulating rod (Fig. 23 – 4). The leaves fall a bit, because we have taken some charge away. The charge is shared between the electroscope and the ball. Try again with a bigger ball. The leaves fall much farther, because the bigger ball takes a bigger share of charge. The same thing happens if we join the electroscope knob to a big ball with a wire. Some of the charge runs along the conducting wire to the big ball. With a very big ball, practically all the charge goes to the ball, and the electroscope is left with so little charge that the leaves fall completely. We can use the largest available ball of all, the earth. When we connect a charged electroscope to the earth, practically all the charge runs away—there is far too little left to notice. (This process of sharing charge with the earth is often called "grounding.")*

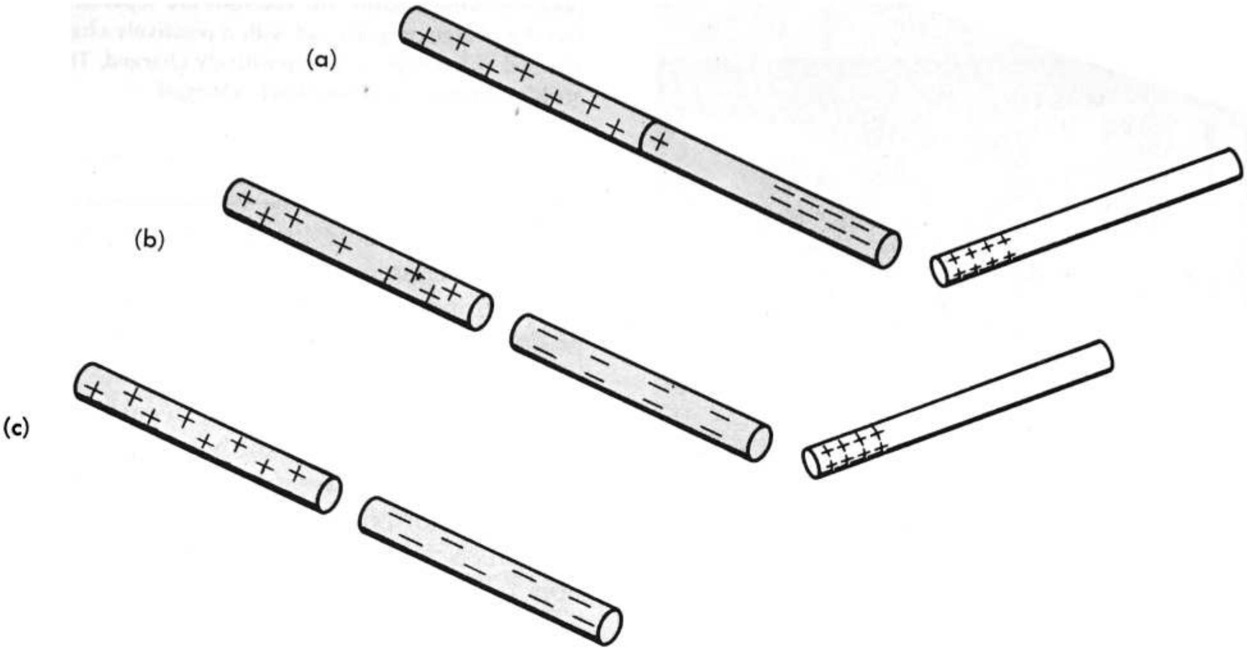
Although you are not as large as the earth, you too are a conducting ball. If you charge the electroscope and then touch the knob with a finger, the leaves collapse. If there was initially an excess of negative particles on the metal rod and leaves, some of these particles must have gone onto your body. They may have passed through your body to the ground. But even if your shoes are insulators, the electroscope leaves will fall. In this case, you are sharing the charge with the small electroscope.

If the electroscope is positively charged, you will still discharge it by touching the knob. Some negative particles must have gone from you to the electroscope, neutralizing the positive charge on the leaves. In either case, from these experiments we conclude that the human body is an electrical conductor.

We can determine what other substances are conductors. For example, suppose we charge the electroscope again and touch the knob with a stick of graphite, such as the lead from a pencil (Fig. 23 – 5). The leaves fall immediately, showing that graphite is also a conductor. On the other hand, the leaves of a charged electroscope do not move if we touch the knob with uncharged glass, hard rubber, porcelain, or plastics. These substances are electric insulators. If we touch the knob with a wooden match, the leaves fall, but very slowly. Apparently in wood some electric particles can move, but not as freely as in metals; wood offers a much greater "resistance" than metals to the motion of electric particles.

---

*The earth has some patches of nonconductor—desert sand, some rocks, etc.—but underground there is a good enough conducting material so that the charge is quickly shared all over the earth. A metal pipe, such as part of the water system, is often used to make sure of the connection to the conducting region underground.



23–6 Electrostatic induction. (a) Two metal rods are in contact; a positively charged object is nearby. Negative particles are attracted to the right, leaving a positive charge to the left. (b) The rods are separated with the charged body still nearby. The right-hand one has a negative charge and the left-hand one positive. (c) The rods retain their charges even when the charged object is taken away.

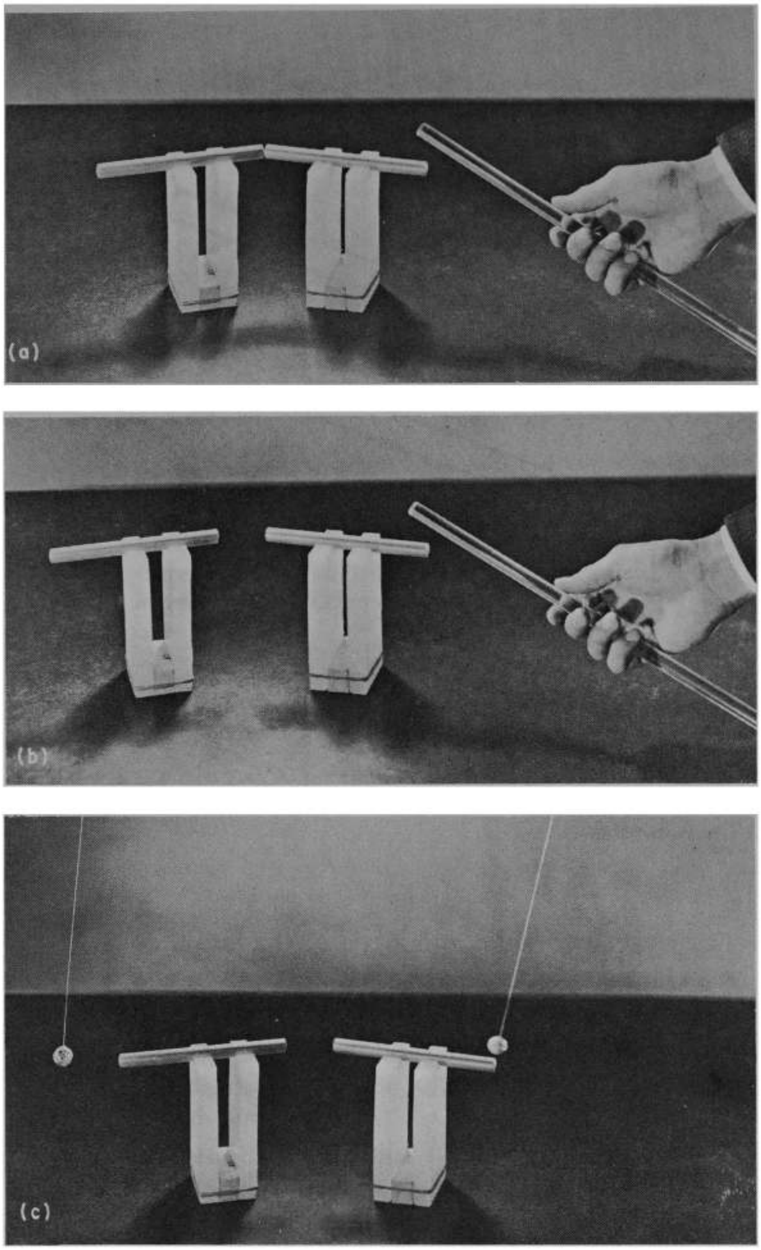
23–5 ELECTROSTATIC INDUCTION

We do not need to touch a conductor in order to move charges around on it. Suppose we have two metal cylinders on insulating stands. We put them in contact to form one long conductor, as in Fig. 23 – 6 and 23 – 7. Then we bring a positively charged glass rod near one end of the conductor. The positive charge on the glass rod will attract negative charges on the conductor and repel positive charges. As a result, the near end of the conductor acquires an excess of negative electric particles and the far end is left positive. Now we separate the two metal cylinders by pulling the insulating stands apart while the positively charged glass is nearby. The near cylinder should have a net negative charge and the far one a net positive charge. We can verify this conclusion by taking the glass away and bringing up a light, positively charged ball on a thread. Then we see that the near cylinder attracts the positively charged ball and the far one repels it.

The separation of positive and negative charges on a conductor induced by the presence of a charged object nearby is called *electrostatic induction;* and the local excesses of positive and negative charge which accumulate in different regions of the conductor are called *induced charges.*

Electrostatic induction allows us to use an electroscope to detect the presence of a charge without transferring it to the electroscope. In fact, we can even determine the sign of the charge. To do this, we first charge the electroscope, let us say positively. Then we bring the unknown charge near the knob. If the charge is positive, it induces a negative charge on the knob and thus increases the positive charge on the leaves. Consequently, the leaves diverge farther [Fig. 23 – 8 (a)]. If, on the other hand, the unknown charge is negative, the induction goes just the other way. The net positive charge on the leaves decreases, and the leaves fall [Fig. 23 – 8 (b)].

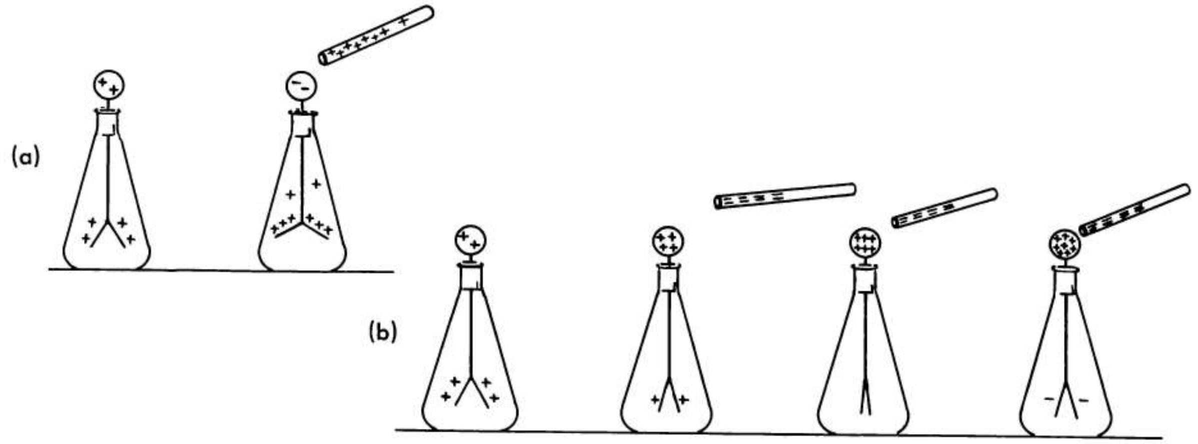
Electrostatic induction also enables us to understand the force of attraction exerted by an electrified body on a neutral conductor. Suppose that the electrified body *A* is positively



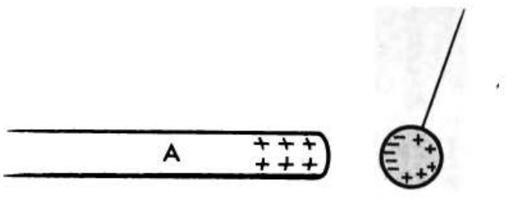
23–7 **Electrostatic induction. (a) Metal rods in contact, positive charge nearby. (b) The rods are separated. (c) We test the charge on each rod with a positively charged ball. The rod at left repels—it is positively charged. The one at right attracts—it is negatively charged.**

charged (Fig. 23 – 9). It will induce a negative charge on the nearest part of the conductor and a positive charge on the part farthest away. The negative induced charge is attracted toward the positive charge on *A*, and the positive induced charge is repelled by it. However, the repulsion is weaker than the attraction, because the positive induced charge is farther from *A* than the negative one. As a result, there is a net force on the conductor pulling it toward *A*.

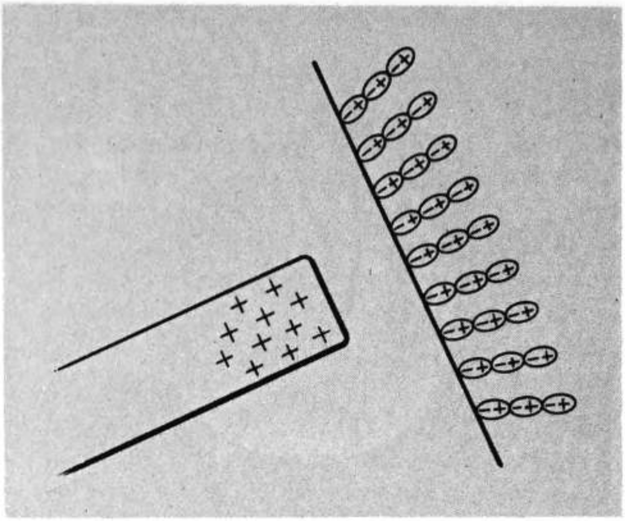
There is a similar but smaller force of attraction between a neutral insulator and a neighboring charge. In an insulator, neither the positive nor the negative particles can wander freely from their atoms, but they can be pushed or pulled over short distances. For example, we may imagine that an insulator is made of positively charged particles nailed down so that they remain in place in the insulator, and negative particles held near the positive ones by springy forces. The springy forces prevent the negative particles from moving any great distance. But when we bring a positively charged body near the insulator, the negative particles are attracted toward it. Consequently, they move a short distance away from the positive particles toward the attracting body (Fig. 23 – 10). Because the negative particles are slightly nearer the positively charged body than are



23–8 **Using an electroscope to detect a charge and determine its sign. In (a) we bring a positive charge near a positively charged electroscope. The leaves diverge still farther. In the series of pictures (b) we move a negative charge nearer and nearer to the knob of a positively charged electroscope. More and more negative charges are repelled toward the leaves, and they fall. Finally, if a big negative charge gets very close, we may push so many minus charges on the leaves that they are given a net negative charge. They then diverge again.**



23–9 **Attraction of a neutral conductor. When a neutral conductor is near a positively charged body, negative electric particles are attracted to the end near the positive charge. A positive charge is left on the rest of the conductor. Since the induced negative charge is nearer the positive body, there is a net force of attraction.**

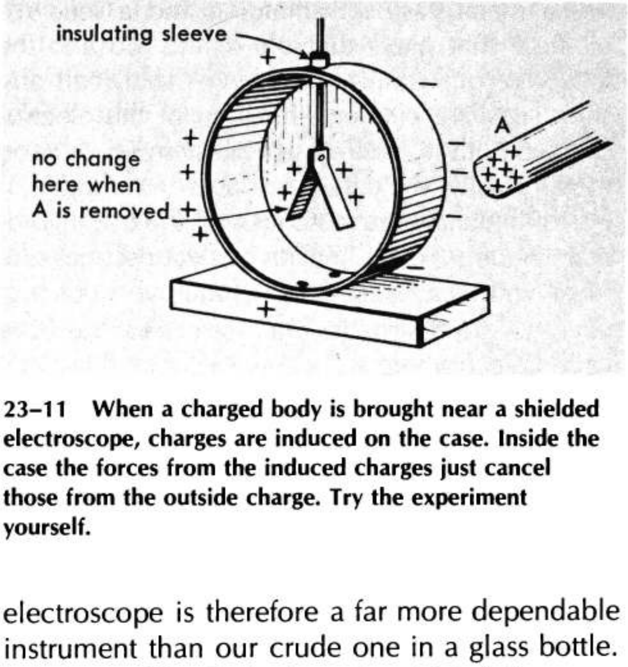


23–10 **Attraction of a neutral insulator. A positively charged object is near an uncharged insulator. There are no electric particles free to move through the insulator, but the positive and negative particles in each molecule can be displaced slightly. The negative particles are thus all slightly nearer the plus charge, giving again a net force of attraction. Note that in fact the molecules are extremely small and so are the charges on them. This diagram is far from being a realistic picture.**

their positive counterparts, the force of attraction on them is slightly greater than the repulsion exerted on the positive particles. As a result, there is a net force of attraction between the insulator and the positively charged body.

23–6 IMPROVED ELECTROSCOPES AND ELECTROMETERS

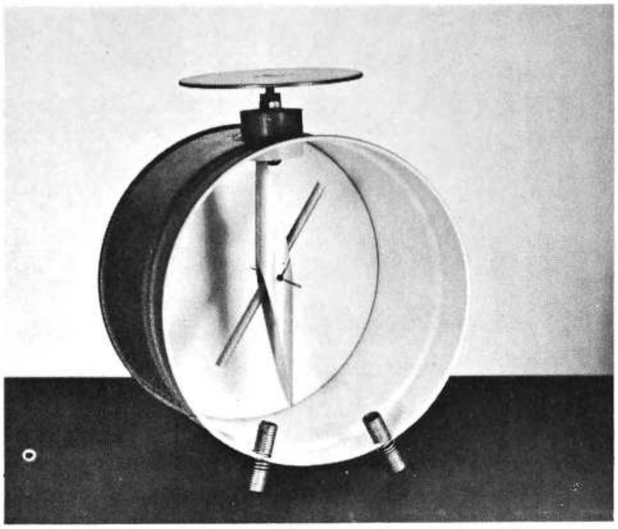
When the gold leaves of an electroscope hang in a glass bottle (as in Fig. 23 – 3), the electroscope is sensitive to charges anywhere nearby. Such charges exert forces on the leaves, both directly (when the leaves are charged) and by electrostatic induction. For many purposes, however, we would like an electroscope to respond only to charges on the leaves or near the knob.



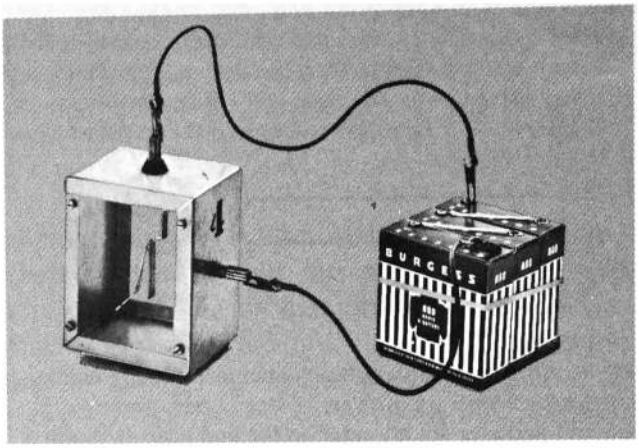
23–11 **When a charged body is brought near a shielded electroscope, charges are induced on the case. Inside the case the forces from the induced charges just cancel those from the outside charge. Try the experiment yourself.**

We can cut out unwanted electric effects by replacing the glass bottle with a metal case and using an insulating sleeve to keep the case and the leaves electrically separated (Fig. 23 – 11). Suppose we bring a positively charged body, *A*, near an electroscope with a metal case. The charge on *A* induces negative charges on the near side of the metal case, and positive charges are left on the opposite side. We find experimentally that if the electroscope is completely surrounded by a metal shield, the leaves are entirely uninfluenced by the presence or absence of charges such as *A* outside the shield. The induced charges exert forces on the leaves, and apparently these forces just balance the force exerted by the positive charge of *A*. A shielded electroscope is therefore a far more dependable instrument than our crude one in a glass bottle. It is insensitive to irrelevant outside influences. You can make a shielded electroscope like the one in Fig. 23 – 12.

Sometimes we wish to detect charges smaller than those which cause an ordinary gold-leaf electroscope to respond noticeably. By changing the design — making the moving part of light



23-12 A simple electroscope shielded by a metal case — a coffee can. Instead of thin metal leaves, the sensitive element is a metal-coated drinking straw hung by a pin that fits through two holes in the frame.



23-13 Charging an electroscope with batteries. The three radio "B" batteries required to produce this deflection of the leaf are equivalent to 180 flashlight cells in series. The effect of a single flashlight cell would not be detectable.

conducting thread, changing the structure of the rod, and making the dimensions smaller — we can build electroscopes which are far more sensitive. Each of these sensitive instruments has a metal case for shielding and a sensitive element that plays the role of the rod and the leaves of our simple model. We shall call any such sensitive element the stem of the electroscope. With a well-designed stem in a good metal shield, we can measure a thousandth or less of the charge detectable with the usual gold-leaf model. Some sensitive electroscopes are fitted with scales so that quantitative studies of charge can be made. Such electroscopes are called *electrometers.*

In designing an electroscope, we took advantage of metal shielding. The observation that a metal shield separates the electrical universe into two parts really does far more than allow us to design a better instrument. The observed cancellation of the electric effects of a charge outside the metal case gives us information about the electrical forces. With careful reasoning we can show that complete cancellation occurs only in the case of forces which vary as the inverse square of the distance.* The Coulomb experiment which we shall discuss in the next chapter shows directly that electrical forces are of this kind.

23-7 BATTERIES

Rubbing two things together (such as a plastic rod and a woolen cloth) is not the only method of separating electrically positive from electrically negative particles. Batteries (and other electric generators) are really machines for separating charges.

A good gold-leaf electroscope will detect the charge separation produced by several batteries (Fig. 23 – 13). The charge that a battery puts on an electroscope is the same kind of electricity that we put on from a rubbed glass rod or from any other charged body. When the plus terminal of a battery is connected to the knob of an electroscope and the minus terminal to the metal case, the knob gets charged with the same kind of charge that we get from rubbed glass — positive charge. When the battery is connected the other way around, the knob becomes negatively charged; and these negative charges can be canceled off by just a small part of the positive charge from rubbed glass.

A battery, then, is a complicated chemical mechanism which shoves plus charges to the positive terminal and minus charges to the negative terminal, despite the forces of electrical attraction which tend to pull these charges together. When the battery is connected to the metal case and to the stem of an electroscope, the case and stem are given opposite charges. The charges quickly accumulate, until after a brief time the electric forces shoving the charges back toward the battery compensate the forces exerted by the chemical action driving them out of the battery.

---

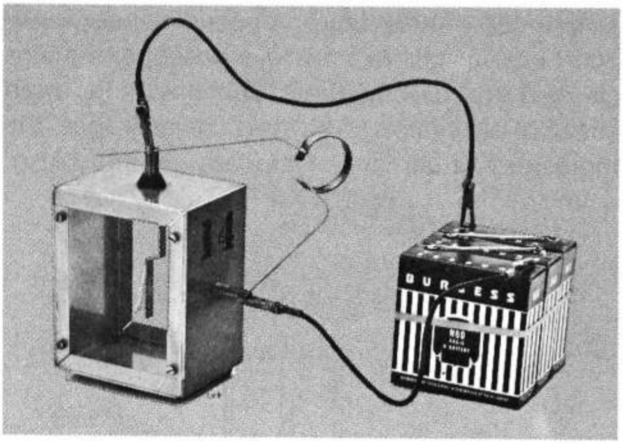
*We shall not go through this reasoning here. An example of it is given in the PSSC film on Coulomb's law, and a related discussion is given in the box in Section 24 – 3.

23–8 ELECTRIC CURRENTS

As we have just seen, we can charge an electroscope by momentarily connecting the stem and the metal case to the terminals of a battery. When we disconnect the battery the electroscope stays charged. But if we then join the stem to the metal case with a metal wire, the electroscope quickly discharges. Electric particles must have moved through the wire between the stem and the case. We shall assume that these are negative particles.

Now suppose that we leave the battery permanently connected to the electroscope. Then if we connect the stem to the case with a poorly conducting wire (Fig. 23 – 14), the electroscope remains partially charged.* There is now a continuous flow of electric particles in the wire. As these negative particles reach the positive battery terminal, they cancel some of the positive charge there; as they leave the negative battery terminal, they drain negative charge from it. The battery, on the other hand, keeps providing a supply of charges at the terminals. Chemical changes continuously drive negative particles across inside the battery from the positive to the negative terminal (or positive particles in the other direction). The rate at which the battery moves these charges is controlled by the rate at which the charges outside the battery flow through the wire and cancel.

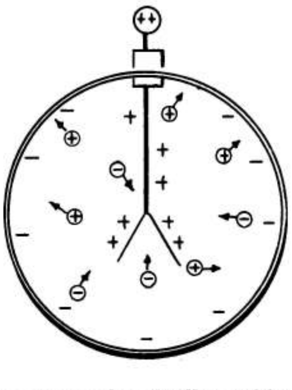
A flow of electric particles in a conductor is called an *electric current.* In particular, the steady flow imposed by a battery is called a *steady current.* Electric currents have many different effects. For example, a metal wire carrying a big current heats up. We can build instruments to detect these effects and thus measure electric currents. In the next few sections and in Chapter 25 there is much more discussion of electric current and its measurement.



23–14 An electroscope connected to a battery stays charged even though the stem and the case are joined by a high-resistance wire. Since the resistance wire is very fine and has to be very long, it is wound up in a coil around the circular form. If the battery is disconnected, the electroscope discharges.

23–9 CONDUCTIVITY OF GASES; IONIZATION

Let us charge an electroscope, for example by momentarily connecting the stem to the positive terminal and the case to the negative terminal of a battery. The electroscope, of course, is full of air. What would happen if some of the air molecules were electrically positive and some electrically negative? The positive molecules would be attracted toward the negatively charged case and the negative molecules toward the positively charged stem. The arrival of the electrically negative molecules would neutralize the excess of positive particles present on the stem, and the electroscope would gradually discharge (Fig. 23 – 15).

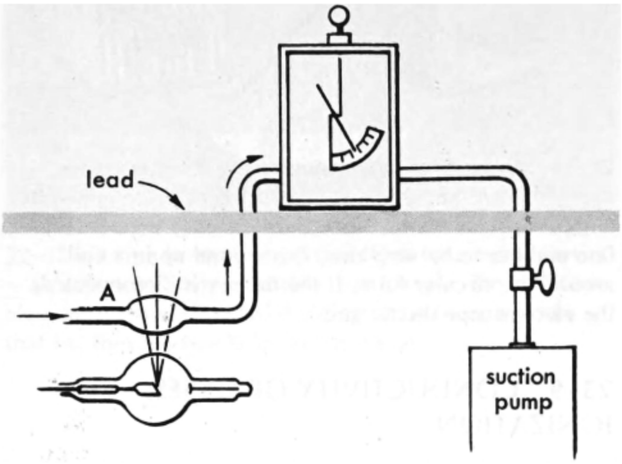


23–15 If air contains electrically positive and negative particles, the charge on an electroscope will be rapidly canceled. Normally this is not true; but air can be ionized.

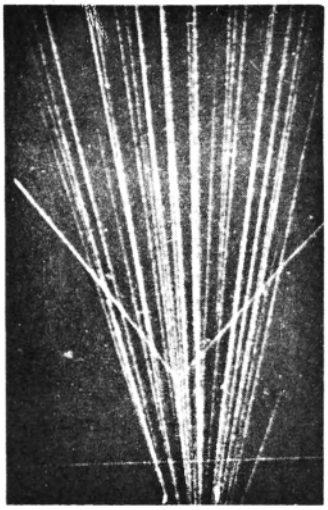
---

*A long, thin wire of a nickel-chromium alloy will do.

Nothing of this kind happens. Under ordinary circumstances, the electroscope remains charged for a long time. The gas in it is an effective insulator. And this means that the molecules of air are *electrically neutral*. Also, if we fill the electroscope with other gases, such as carbon dioxide, helium, argon, etc., we always find the same result, even when the gases are monatomic. Thus complete molecules, including those that are single atoms, are electrically neutral.



23–16 X rays ionize the air at A. When this air with the ions reaches the electroscope, the leaves fall.



23–17 A cloud-chamber photograph showing the paths of alpha particles in a chamber containing helium. (From P. M. S. Blackett, in Proceedings of the Royal Society, 107A, 349; 1925.)

But now let us bring some radioactive material near the charged electroscope, or let us direct a beam of X rays toward it. The electroscope will gradually discharge. Furthermore, if we remove the radioactive material (or shut off the X-ray beam) before the electroscope has lost all its charge, it will stop discharging.

We can interpret these results by assuming that the X rays, or the radiations from the radioative material, break up the gas molecules into electrically charged fragments. Some of these fragments may be identical with the ultimate building blocks of matter that we have been talking about. Others may be composed of serveral such particles. Some may be made of a charged fragment of a molecule that has attached itself to a neutral molecule, thus producing a heavier charged unit. The charged molecular fragments or more complex units are called *ions;* and a gas containing ions is said to be *ionized.*

While an ordinary gas is an insulator, an ionized gas behaves like a conductor. A positively charged object immersed in an ionized gas attracts the negative ions; a negatively charged object attracts the positive ions. In either case the ions, as they come in contact with the object, gradually neutralize the original charge.

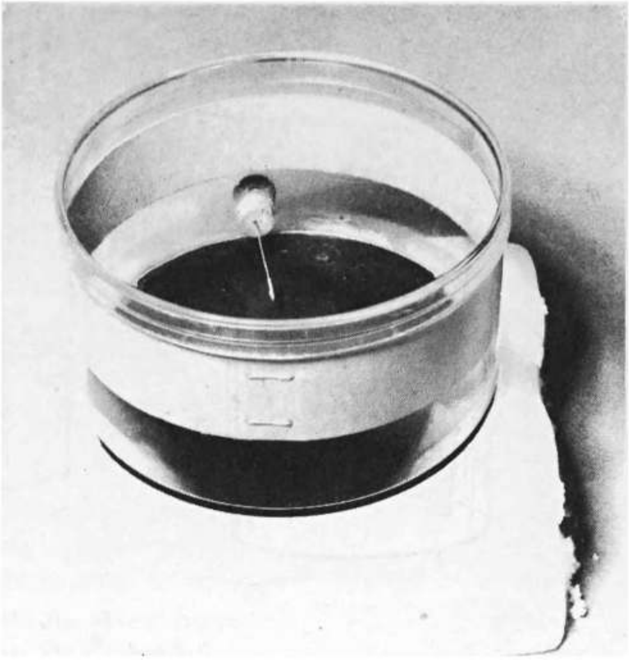
There are many experiments we can do to show that the conductivity of gas results from ionization. For example, we can make sure that the discharge of an electroscope placed in a beam of X rays is not due to a direct effect of the X rays on the stem of the electroscope. The experiment illustrated in Fig. 23 – 16 shows that the X rays are acting on the gas. In this experiment we pump air into the electroscope through a tube, part of which can be irradiated. We charge the electroscope and start the pump. Then we turn on the X rays. The electroscope begins to discharge, because ions formed by the X-ray beam are carried into the electroscope case by the moving air. With the X-ray machine still running, we stop the pump. The flow of air stops, and the ions formed in the tube no longer reach the electroscope. The gradual discharge of the electroscope also stops.

At the beginning of this section we said that gases usually are good insulators. We might have said that in the absence of ionizing radiation they are perfect insulators. As a matter of fact, however, a small amount of radiation is present everywhere. Most materials around us contain minute traces of radioactive substances. And although we can eliminate radiation from materials nearby, cosmic rays will make some ions. Even exceedingly thick shields made of materials free from radioactivity are incomplete protection against these penetrating radiations that originate outside of our atmosphere. Thus, gases are always slightly ionized and therefore slightly conducting. This conductivity, however, is so small that it can only be detected with very delicate apparatus.

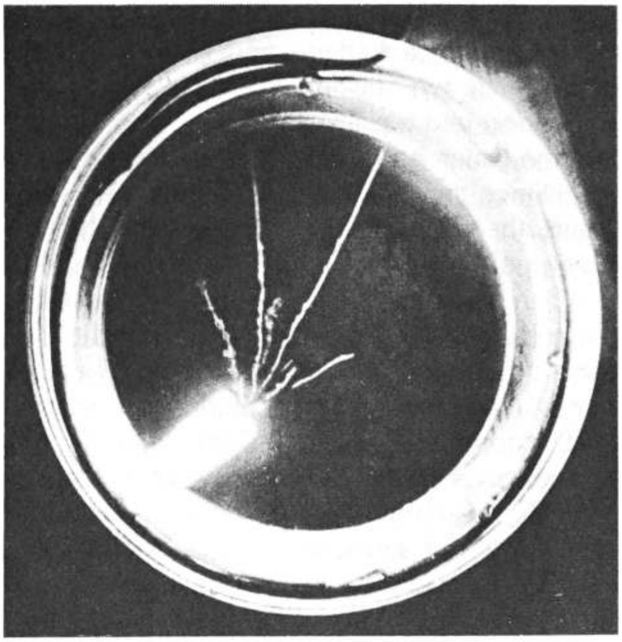
23–10 THE CLOUD CHAMBER

A *cloud chamber* enables us to "see" the paths of submicroscopic charged particles in gases (Fig. 23 – 17). The particles leave a trail of positive and negative ions; and the operation of the cloud chamber is based upon the observed fact that vapors condense more easily around such ions than in open space.

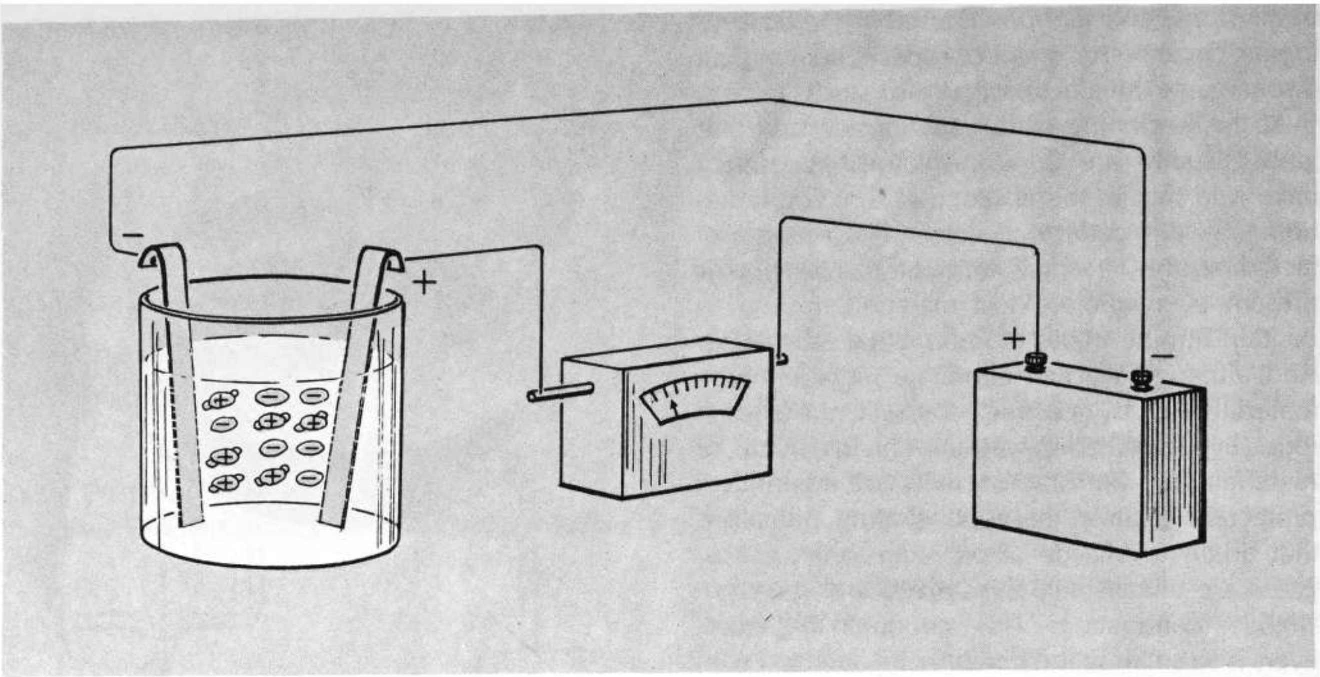
A type of cloud chamber called a diffusion chamber is shown in Fig. 23 – 18. A simple version of it consists of a shallow cylindrical box, the bottom of which is kept cold while the transparent top is kept at a higher temperature. On the wall of the chamber there is a wick of cloth or of blotting paper soaked in alcohol. The alcohol continuously evaporates from the wick and condenses on the cold floor. Just above the floor there is a region where the alcohol vapor does not condense spontaneously but will condense around any ion formed. When a charged particle traverses this region, its track becomes visible. This cloud chamber is cheap to make, and it runs continuously. With it we can see the passage of an occasional charged cosmic ray. Also, by putting a bit of radioactive material in the chamber we can see the tracks of the charged fragments emitted in radioactive decay, as in Fig. 23 – 19.



23–18 A simple cloud chamber. The chamber itself is a cylindrical plastic box which rests on a large block of Dry Ice. A dark felt band, soaked in methanol, encircles the inside of the top of the chamber. A tiny radioactive source is placed in the eye of the needle projecting into the chamber. The needle is supported by a cork stopper inserted in a hole in the wall of the cylindrical box. The bottom of the chamber is painted black so that the white fog tracks will be plainly visible when viewed from above and illuminated from the side.



23–19 Tracks of alpha particles in the diffusion cloud chamber.



23–20 Apparatus for studying electrical conduction in a liquid. When salt is dissolved in the water, the meter shows a current. The charge is carried through the solution by positive and negative ions.

23–11 CONDUCTIVITY OF SOLUTIONS

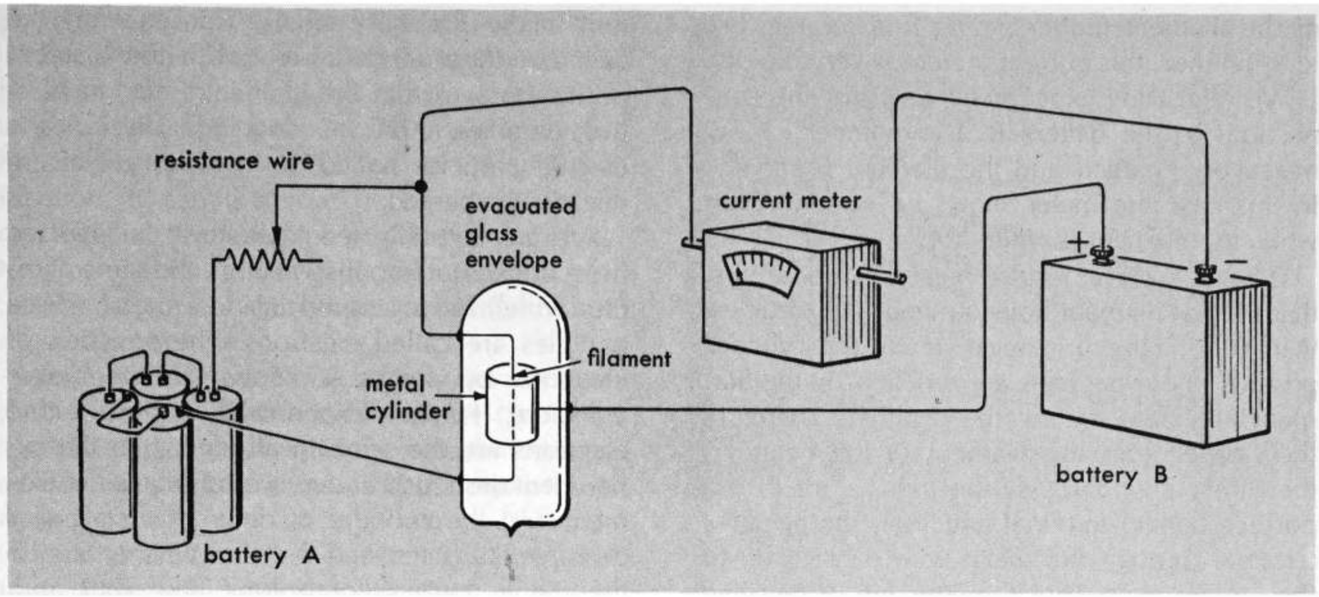
Some liquids are electrical conductors; others are insulators. The behavior of water and of water solutions is particularly interesting. In order to study this behavior experimentally we shall use the arrangement shown in Fig. 23 – 20. In series we connect a metal plate, an electric current meter, a battery, and a second metal plate. Then we stick the metal plates (called *electrodes*) into a glass of water.

If the water is pure, the meter shows no current. Since the molecules in liquids are free to move, the absence of current proves that the water molecules are electrically neutral.

When we dissolve some table salt in the water, the meter shows a current in the circuit. The same thing happens with many other chemical compounds such as hydrochloric acid, sulfuric acid, copper sulfate, potassium chloride, etc. We can explain the conductivity of these solutions by assuming that the table salt (or the other dissolved substance) breaks up into positively and negatively charged fragments, which we shall again call positive and negative *ions*. This explanation is much like the explanation of electric conduction in gases. In liquids, however, unlike gases, no ionizing agent outside the system is needed. The ions seem to be formed after a spontaneous breakup of the dissolved substance.

We get convincing support for our explanation of the electrical conductivity of solutions by observing what happens at the electrodes. If we have a copper salt (for instance copper sulfate) dissolved in the water, we find that when there is an electric current the negative electrode gradually becomes coated with a film of copper. If, instead, we fill the cell with a solution of a silver salt (such as silver nitrate), we find that the negative electrode becomes coated with a film of silver. This suggests that when copper sulfate or silver nitrate breaks up in solution, the *metal* atoms are carried by the *positive* ions. The motion of these ions in the electric current involves a detectable transport of matter.

The phenomenon described above is known as *electrolysis*. It is commonly used in electroplating and in many other manufacturing processes, such as refining copper, producing hydrogen gas, and producing aluminum. In Section 25 – 3 we shall see how to use electrolysis to measure electric current.



23–21 **Apparatus for investigating electrical conduction in metals. The electric current from battery A heats the filament. We can raise the temperature of the filament by removing resistance wire from the circuit of battery A. The current meter connected to battery B shows when charge flows through the space between the filament and the cylinder.**

23–12 ELECTRONS IN METALS

Metals are the commonest kind of conductor. What kind of electric particles are free to move in them, making them conduct? Are these particles positive or negative or both? No single experiment is capable of answering this question in such a clear-cut manner as in the case of conducting gases and liquids. However, one observation gives us a very suggestive clue; and together with other pieces of circumstantial evidence it provides an answer.

It is natural to think that the electric particles in metals are in a state of thermal agitation like the molecules of a gas (see Chapter 20). Yet experiments show that a charged piece of metal in a vacuum holds its charge indefinitely. Consequently, we know that the electric particles cannot ordinarily escape from the metal. And we conclude that, when they are near the surface of the metal, the moving particles experience forces that prevent their escape by attracting them back into the metal.

Now suppose we heat the metal. As the temperature increases, the speeds of the particles must increase, just as the speeds of gas molecules do. At high enough temperatures an appreciable fraction of the particles should have sufficient speed to escape. Like rockets moving fast enough to escape from the gravitational attraction of the earth, these high-speed particles should not fall back into the metal. They should get out beyond the reach of the attractive forces.

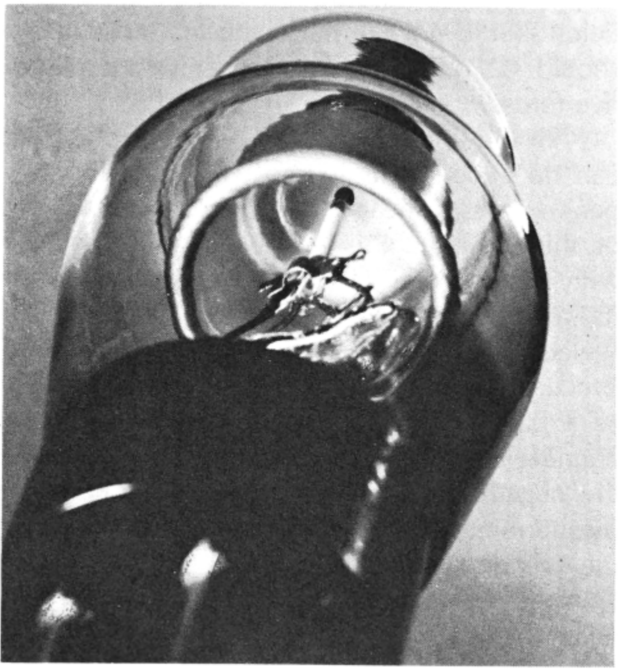
With the apparatus shown in Fig. 23 – 21 we can do experiments to see what actually happens when we heat a metal. The essentail part of this equipment is a hollow metal cylinder with a thin metal wire filament stretched along its axis. The cylinder and the filament are sealed in a glass vessel, which is thoroughly evacuated. To heat the filament we connect it to a battery marked *A* in the figure. We also connect the filament through a current meter to the negative terminal of a battery *B*. The positive terminal of the battery *B* is connected to the cylinder.

When the filament is at ordinary temperatures, there is no current in the *B* battery circuit. But if we gradually raise the temperature of the filament, the meter will eventually register a current, indicating that electric particles pass through the empty space between the filament and the cylinder. As we raise the temperature of the filament further, going from orange hot to white hot, this current increases very rapidly.

We repeat the experiment, reversing the connections to the battery *B*. The cylinder is now *negatively* charged and the filament *positively*. In this case the meter shows no current, even when the filament is white hot.

What do these results mean? Suppose first that the hot filament boils out negative particles. When the filament is negative and the cylinder positive, these particles are repelled by the filament and attracted by the cylinder. Therefore they travel from the filament to the cylinder; they then flow through the metal wire to the positive battery terminal and from the negative terminal through the metal wire back again to the hot filament. Thus the emission of negative particles from the filament effectively bridges the gap between the filament and the cylinder so that a steady current is observed in the circuit.

Suppose, on the other hand, that the filament emits positive particles. They would bridge the gap when the filament is positive and the cylinder negative; and we would then observe a current in the *B* battery circuit. Consequently the fact that there is no current with the filament positive proves that the filament emits no positive particles. And we conclude that all the electric particles boiled out of a hot metal are negatively charged.



23–22 **A photograph of a commercial diode. The white metal tube in the center is heated internally and emits electrons. When it is charged negatively and the outer cylinder is positive, the electrons move through the space between.**

We are naturally led to believe that we see here, boiled out into the vacuum, the same particles which move around inside a metal. These particles are called *electrons*. The emission of electrons by metals is known as *thermionic emission*. Further experiments indicate that electrons are the same in all metals. In the experiment described above, we can make the filament and the cylinder of different metals—for example, tungsten and nickel. We then run a thermionic current for a long time (but at a temperature at which tungsten does not evaporate appreciably). We find no trace of the filament's metal on the cylinder, no matter how long electrons have been flowing across. There is no change in the chemical composition of either the cylinder or the filament. The electrons that have come out of the tungsten must be identical with the electrons already inside the nickel.

This conclusion is supported by similar evidence from metals in contact. We can leave a nickel wire connected across the brass terminals of a battery and let a current run for days in succession without noticing any change in the wire itself or in the battery terminals. A great many electrons have gone from one metal to the other across the contact, but no change takes place in the brass or the nickel. The electrons of nickel and brass are thus indistinguishable.

Finally, we conclude that electrons must be among the ultimate building blocks of all metals and in fact of all atoms. Today the correctness of this conclusion is beyond doubt. We shall meet more of the evidence for it in later chapters.

23–13 DIODES, ELECTRON GUNS, CATHODE-RAY OSCILLOSCOPES

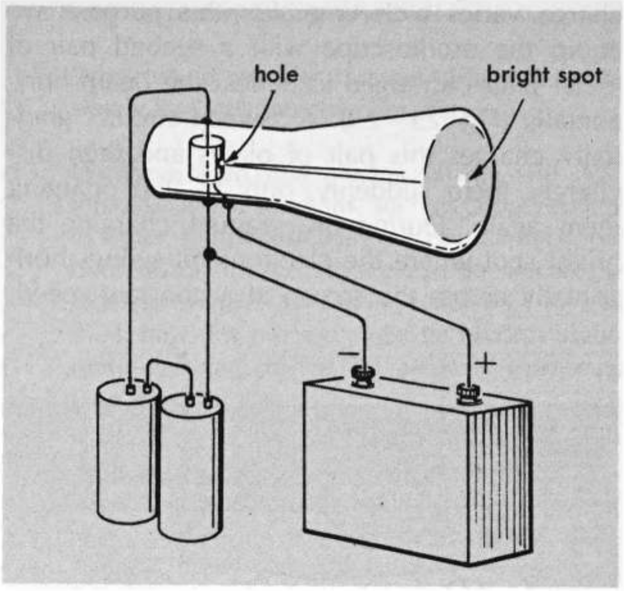
Thermionic emission has many practical applications; one of the most important is in radio tubes. The device we used to demonstrate thermionic emission is a simple type of radio tube called a *diode* (Fig. 23 – 22). (In the language of electronics the hot wire from which the electrons are emitted is called the *cathode*, and the electrode to which the electrons travel is called the *plate*.) In a diode, electrons flow from the cathode to the plate, but not in the opposite direction. It can therefore be used as a valve (or "rectifier" or "detector") to change a current that flows alternately in each direction in a circuit into a current that flows intermittently but in one direction only.

We can also use a diode to make an electron gun by drilling a hole in the plate. (See Fig. 23 – 23.) Although most of the electrons driven across from the hot filament hit the plate, those moving straight toward the hole go through and travel on until they meet the wall of the tube. If the wall opposite the electron gun is covered with fluorescent paint, a bright spot appears where the electrons hit.

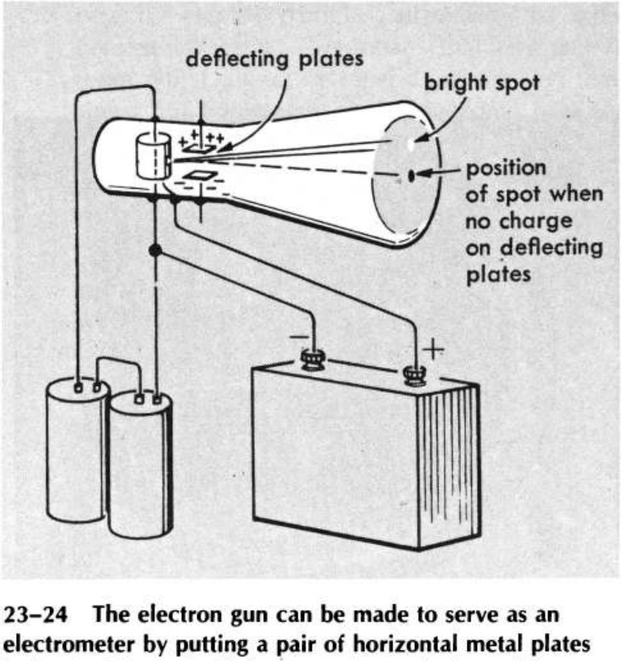
With one more modification, the tube we have described becomes a cathode-ray oscilloscope. We put two small horizontal metal plates in the tube so that the electron beam passes between them (Fig. 23 – 24). If one plate is charged positive and the other negative, the electrons are shoved vertically toward the positive plate and hit the fluorescent screen at a new place. Even a single flashlight battery connected across the deflecting plates produces a visible deflection of the spot. Consequently, we can use this oscilloscope as an electrometer. And if we couple it with an amplifier like the one in a radio, it can be a far more sensitive instrument than the electrometers we described previously.

In addition to its sensitivity as an electrometer, the cathode-ray oscilloscope has an even more useful property: the speed of its response. Because of the mass of its moving parts, an ordinary electrometer takes an appreciable time to reach the equilibrium position after it has been charged. The electrons, on the other hand, have astonishingly small mass compared to the lightest mechanical part. Consequently, the deflection of the electron beam occurs practically instantaneously. Thus the cathode-ray oscilloscope is capable of showing very rapid changes of charge, something that ordinary electrometers cannot do.

With the cathode-ray oscilloscope we can even study in detail how a rapidly changing

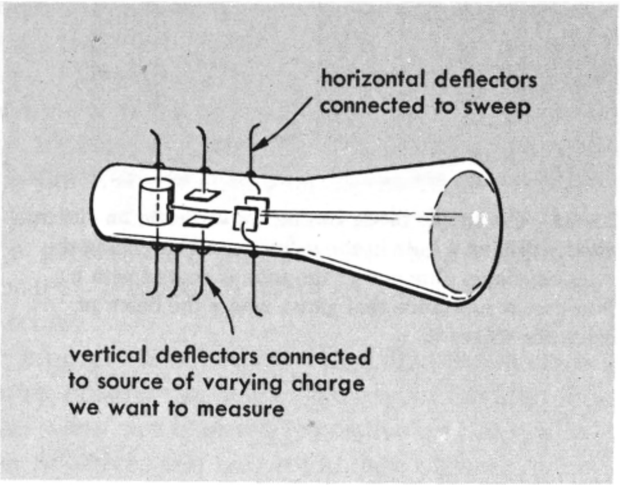


23–23 The simple diode can be converted to an electron gun by drilling a hole in the cylinder and extending the glass envelope. One end of the tube is coated with a fluorescent substance that glows where the beam of electrons strikes it.

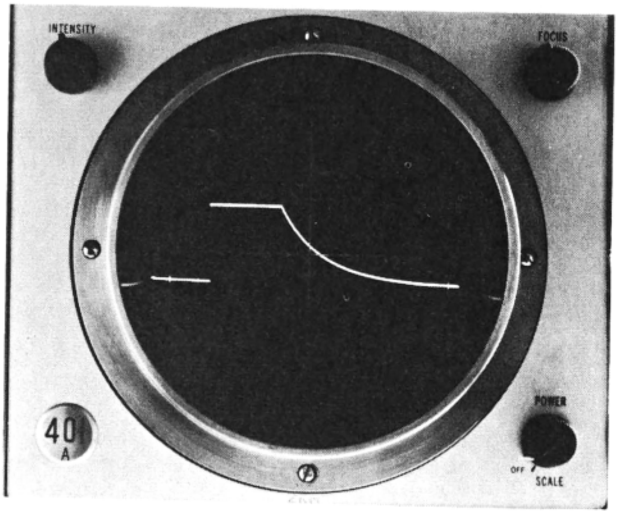


23–24 The electron gun can be made to serve as an electrometer by putting a pair of horizontal metal plates in front of the hole. If, for example, we put a positive charge on the upper plate and a negative charge on the lower, the electron beam will be deflected upward. The movement of the spot away from its center position is a measure of the charge on the plates.

charge varies with time. For this purpose we equip the oscilloscope with a second pair of metal plates arranged to deflect the beam horizontally (Fig. 23 – 25). A "sweep circuit" gradually charges this pair of plates and then discharges them suddenly, only to start charging them again. During the gradual charging the bright spot where the electrons hit swings horizontally across the screen at a constant speed; then, on discharge, it returns quickly to the starting point. At the same time the first pair of plates produces a vertical deflection corresponding to their charges. As a result of the two deflections the spot traces a curve which gives a graphical picture of the way the charge on the vertical plates changes with time (Fig. 23 – 26).


horizontal deflectors connected to sweep
vertical deflectors connected to source of varying charge we want to measure

23–25 An oscilloscope tube has a second pair of plates to give a horizontal deflection. Charging these plates swings the spot across the face of the tube.



23–26 A photograph of a trace on the face of an oscilloscope tube. While the sweep circuit moved the beam from left to right, a battery rapidly charged two metal plates connected to the vertical deflection plates of the tube. A moment later they were allowed to discharge more slowly through a wire of high resistance.

The cathode-ray oscilloscope has many applications. Television and radar tubes are special types of cathode-ray tubes. The pictures they paint are produced by deflecting the electron beam horizontally and vertically over the face of the tube while the strength of the beam is varied to make each little region of the screen glow bright or stay dark in response to signals from the TV camera.

QUESTIONS

1.* You find that object *A* repels object *B*, *A* attracts *C*, and *C* repels *D*. If you know that *D* is positively charged, what kind of charge does *B* have?

2.* Must all the negative charges be removed in order for a rod to become positively charged?

3.* (a) Why can't you electrify a metal rod by rubbing it while holding it in your hand?
(b) What could you do to electrify a metal rod?

4. When the gasoline tank of an airplane is filled, the metal nozzle of the hose is always carefully connected to the metal of the airplane by a wire before the nozzle is inserted in the tank. Be prepared to explain why this procedure is followed, and to describe how it accomplishes its purpose.

5. (a) When you touch a metal object like a door handle on a dry winter day, you are likely to see a spark and feel a definite shock. We usually explain this by saying that we have built up a static charge. How could you tell the sign of the charge you are carrying?
(b) If you accumulate a static charge and then touch the wooden frame of a door, you often find no spark or shock—although there would be if you touched the metal handle. Why?
(c) Sometimes, if you touch the wooden frame and then touch the metal handle, there is no spark or shock in either case—although there would have been one if you had touched the handle first. Suggest an explanation.

6.* What will happen if you touch together the knobs of two identical equally charged electroscopes
   (a) if both are positively charged?
   (b) if both are negatively charged?
   (c) if one is negative and the other positive?

7.* As you bring a charged rod near a negatively charged electroscope, the leaves start to collapse. What is the charge on the rod?

8.* If you move a charged rod toward a positively charged electroscope, the leaves at first collapse and then diverge. What charge is on the rod?

9. Three metal blocks in contact are resting on a plastic tabletop. You place two objects with strong positive charges, one at each end of the line of blocks, close to but not touching the blocks. You then poke the blocks apart with an (uncharged) insulating rod, while the objects with strong positive charges are nearby. Finally you remove the two positively charged objects.

   (a) What charge is now on each block?

   (b) Explain how the blocks acquired these charges by describing the motion of negative particles.

10. You have two metal spheres of the same size mounted on insulating stands, a strip of plastic, and a piece of cotton cloth. Describe what you would do to give the two spheres equal electrical charges (a) of the same sign, (b) of opposite signs.

11. (a) Describe the steps you would take to charge an electroscope positively by induction.

   (b) Using labelled sketches, describe the movement of negative electric particles during the charging process.

12. We think the earth has a slight electric charge. How would you go about testing for it?

13.* What will happen to the electroscope leaves in Fig. 23 – 14 if the resistance wire is left in place and the battery is disconnected?

14.* Why would an electroscope discharge more slowly in a vacuum chamber than in air?

15. We have seen that electric charge will flow through certain solutions, such as copper sulfate in water. In this case there is a deposit of copper built up on the negative electrode. Recall the explanation given for this in Section 23 – 11, and think how this process might be used as a current-measuring device. Be prepared to discuss your ideas in class.

16.* Suppose the entire battery *B* circuit of Fig. 23 – 21 were replaced by a positively charged electroscope with the knob connected to the metal cylinder inside the glass envelope. What would happen when the filament was heated?

17. Suppose we have a large number of identical particles. Any two of them at 10 cm separation repel with a force of 3×10^{-10} newton.

   (a) If one of them is at 10 cm from a group of *n* others, how strongly do you expect it to be repelled?

   (b) Suppose you measure the repulsion and find it is 6×10^{-6} newton; how many particles were there in the group?

18. Suppose we have a large number of positive and negative electric particles, all exactly the same (except for sign). This means that any pair of them either attract or repel depending on their signs, but the magnitude of the force between them is the same as long as the separation distance is the same. For example, the force between any two of them at 10 cm is 3×10^{-10} newton.

   A single positive particle is 10 cm from a group containing *P* positive particles and *N* negative particles.

   (a) How strongly will it be repelled?

   (b) You are unable to count the particles, but you measure the force and find it is 6×10^{-6} newton. What does this tell you about *P* and *N*?

24

Coulomb's Law and the Elementary Electric Charge

SO FAR, our study of electric forces has been qualitative. It indicated that the electric attraction and repulsion between large bodies is due to attraction and repulsion between the subatomic particles of which they are made. Our next task will be to describe experiments that provide quantitative information about these forces.

24–1 FORCE VS. DISTANCE

The way in which the attraction or repulsion between charges depends on distance was established experimentally by the French physicist Charles Coulomb in 1785. In his experiment Coulomb used a torsion balance, like the instrument Cavendish used to study gravitational attraction (see Section 15 – 10). Figure 24 – 1 is a diagram of this kind of apparatus. When the charged sphere *A* is put in place, the electric force it exerts on the charged sphere *B* pushes the horizontal arm around. The arm then comes to rest in a new position with the suspension twisted. The more the suspension is twisted, the greater the force must be. From the angle of twist, therefore, Coulomb could measure the electric force. And by changing this distance between the charged spheres, he measured the force as a function of separation. Because the charges are distributed all over the spheres, the spheres must be far enough apart so that the slight differences in distance and direction do not matter.*

Using both positively and negatively charged spheres, Coulomb showed that the electric force is always inversely proportional to the square of the distance between the charges. He established this result with an accuracy of about 3 percent. Later indirect tests—using the shielding effect of conductors—have shown it with much higher precision. Cavendish attained 1 percent accuracy; and in the latter part of the nineteenth century Maxwell established the exponent 2 (the *square* in the inverse square) to within one part in 40,000. It is now known to one part in 10^9. Note that electric forces and gravitational forces vary with distance in exactly the same manner. We have no ex-

---

*When charged bodies are far apart compared with their dimensions, they are often called *point charges* because the detailed positions of the charges on the bodies are then unimportant.

planation of this similarity, but because of it we can often understand gravitational and electrical effects in the same way.

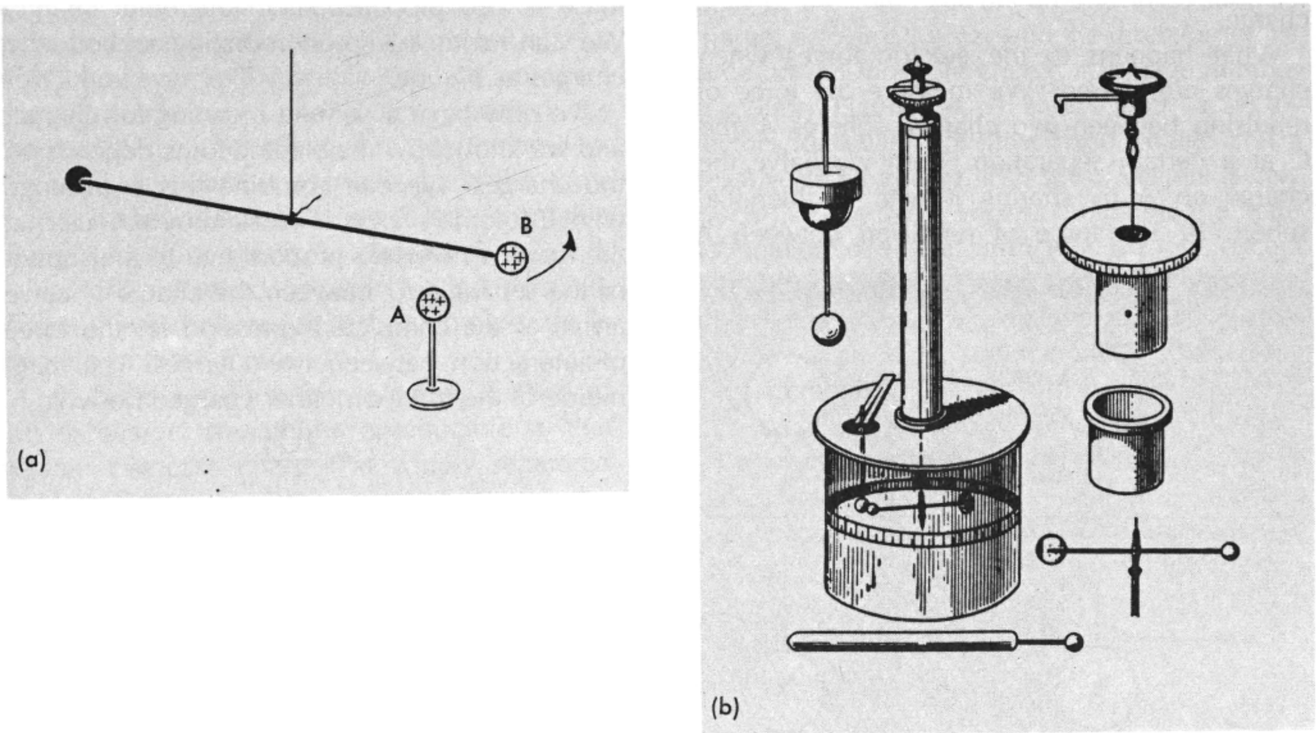
We believe that the force of attraction or repulsion of one charged object on another is the vector sum of all forces that the electric particles of the first object exert upon the electric particles of the second. And in all cases, we find experimentally that the force varies in inverse proportion to the square of the separation. So we naturally conclude that *the force of one individual electric particle on another varies as the inverse square of their separation*. Like the other interaction forces we have studied which are a function of separation, the force of one charged particle on another is equal and opposite to the force exerted on the first particle by the second.

Since 1910 a great many experiments have been done with subatomic particles. We can fire them at one another or at the charged nuclei of atoms. As we shall see later on, these experiments also show that the electric force between the particles varies as the inverse square of the separation.

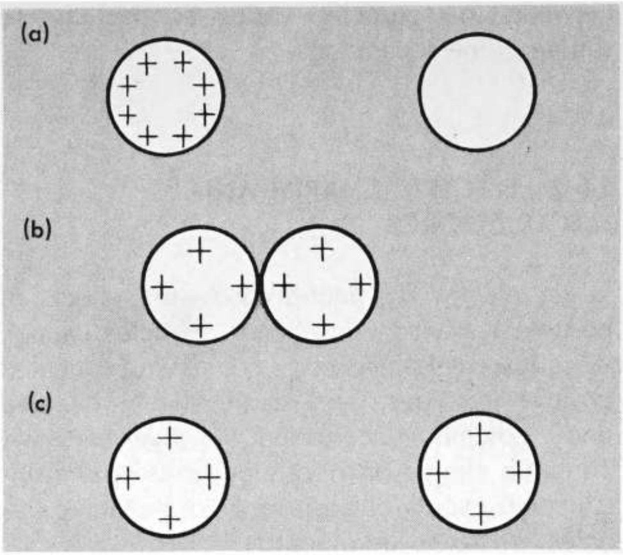
24–2 ELECTRIC CHARGE AND ELECTRIC FORCE

In an electrically neutral body the effects of positive and negative electric particles cancel. A positively charged body contains uncanceled positive particles, and a negatively charged body contains uncanceled negative particles. Thus the charge of a body depends on the uncanceled excess of positive or of negative particles, measured from neutral.

The force between two charged bodies depends on their separation and increases with the excess of positive or of negative electric particles on each body. Just how does the force depend upon the excess of electric particles? To answer this question we need a scheme to divide the excess of particles in a known way—in half, in



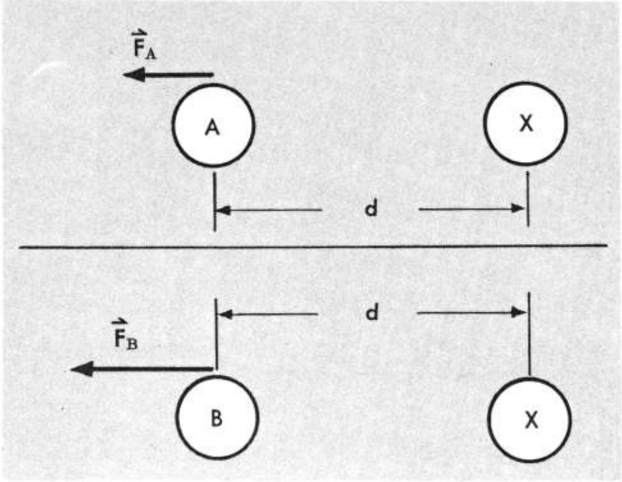
24–1 (a) Schematic diagram of Coulomb's torsion balance. (b) A sketch of Coulomb's apparatus which appeared in his original paper.



24–2 The sharing of electric charge. When a charged sphere is touched to an identical uncharged one, the excess of electric particles divides equally. The final distribution of charge must be symmetrical, as shown in (c).

thirds, etc. Suppose we touch a charged metal sphere with an identical uncharged sphere (Fig. 24 – 2). Then the electric particles will move around until they are shared equally by both spheres. Each sphere will have half the original charge.

What happens to the electric forces when charges are shared? We measure the force of repulsion between two charged spheres *A* and *C* at a certain separation. Then we halve the charge on *A* by sharing it with an identical sphere *B*. The force of repulsion between *A* and *C* (still at the same separation) is *also* cut in half. Furthermore, we get the same force when *A* is replaced by *B*, the identical sphere with which it shared its charge. Apparently, charge and force are proportional, as we might have guessed.



24–3 To compare two charges, A and B, we place them in turn at the same distance from any other charge X, and measure the forces. The ratio of the charges equals the ratio of the forces: $q_A/q_B = F_A/F_B$. What do you think is the ratio of the forces exerted on X?

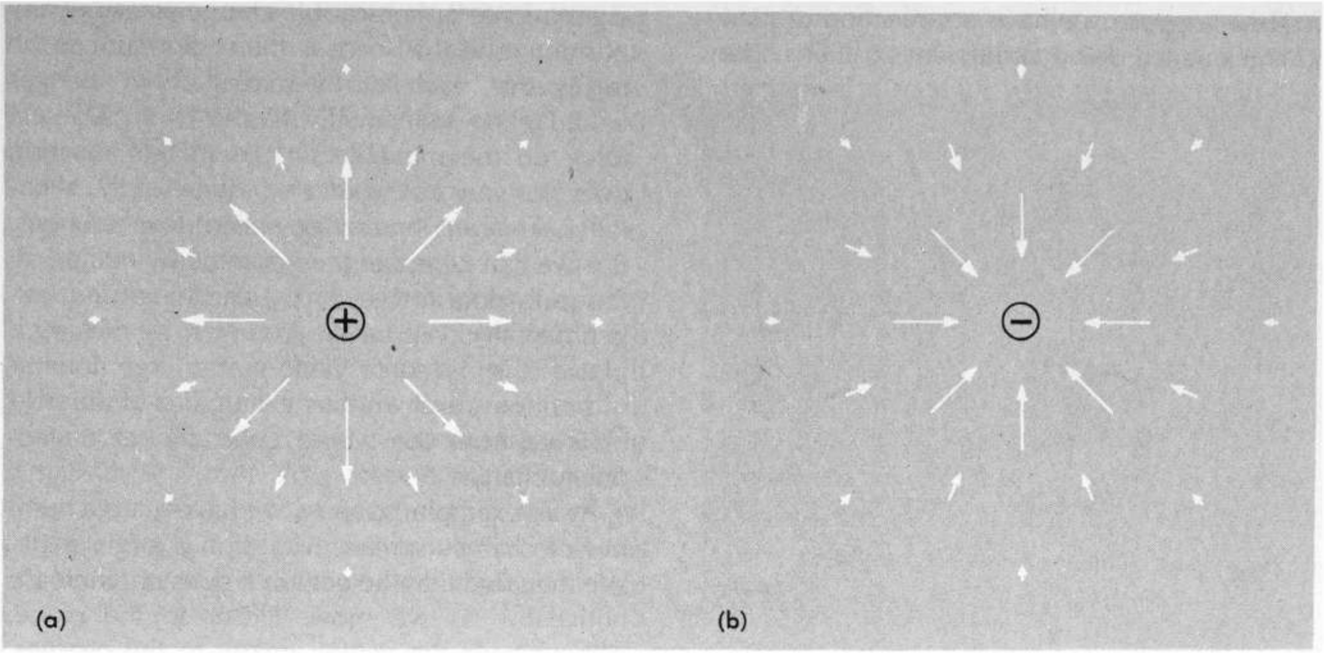
Such experiments give us a way of comparing charges quantitatively. Two charges are equal if they experience equal forces at a given distance from any third charge. One charge is twice another when it experiences twice the force. When a charge is halved by charge sharing, the force exerted on it by a third charge is also halved. In general, charges are compared by the ratio of the forces exerted on them by any other charge at a given distance. This *ratio* does not depend on the magnitude of the "other" charge nor on the distance apart (Fig. 24 – 3). Equivalently, we can compare the ratio of the forces exerted *on* the "other" charge by each of the two charges being compared.

Now let us summarize our knowledge in algebraic language. The electric force on a charge q is proportional to the charge: $F \propto q$. When this force is the force of interaction on the charge q by another small body of charge Q, the force is also proportional to the other charge. We can write this proportionality to both the charges as $F \propto qQ$.

We now have a definite meaning for charge, and we know how the electric force depends on the charges. We can combine this knowledge with Coulomb's experiments. They tell us that the force is inversely proportional to the square of the separation r between the charges. So we arrive at the complete expression for the force of interaction between two charges. The magnitude of the force on either charged body is

$$F = \frac{kQq}{r^2}$$

where the proportionality factor k depends only on the units in which we measure forces, separations, and charges. If Q and q have the same sign, the forces are repulsive and the force on each charged body points outward along the line joining the two bodies. On the other hand, if Q and q have opposite signs, the forces are attractive and the force on each body points inward along the line joining them. We shall call the expression for the magnitude of the force *Coulomb's law*, and the forces themselves the *Coulomb forces*.



24–4 The electric force field exerted on a positive movable charge by (a) a positive fixed charge and (b) a negative fixed charge. It is the set of vectors showing the force on the movable charge at each point in space.

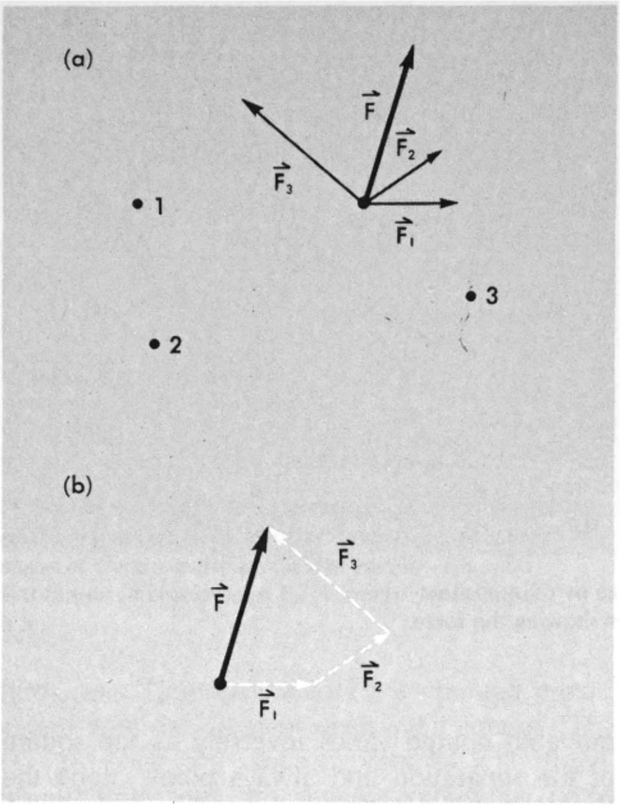
When we first began speaking of electric charge, we gave this word only a qualitative meaning. An electric charge was simply the result of adding or removing some positive or negative electric particles from a neutral object. Now, however, we have learned to compare charges numerically; and we can speak of the charges of the individual particles themselves. We shall say that two particles have the same charge if they exert the same force on another charge placed at the same distance from each. If the forces exerted by the two particles are different, we shall say that their charges are in the ratio of these forces.

One thing is still missing. We need a standard unit of charge, something reproducible so that we can compare charges at widely separated places. For this purpose many different arbitrary units have been used. Fortunately, however—as we shall see in Section 24 – 5—nature itself provides a fundamental unit.

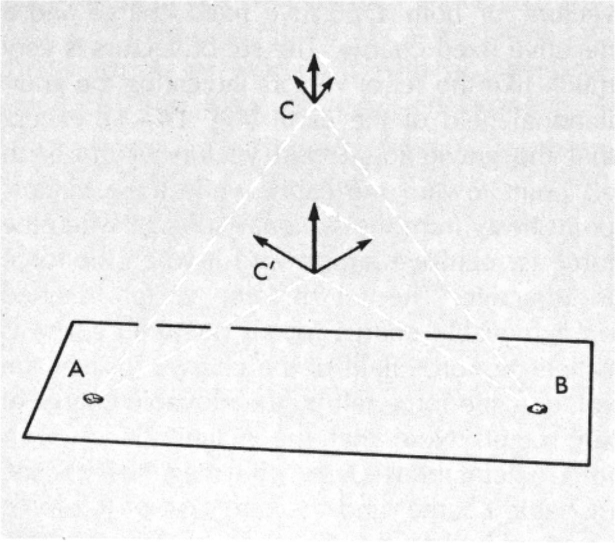
24–3 ELECTRIC FORCE FIELDS

Suppose we move a positive charge to various positions near another small charged body fixed at a particular place. The force exerted on the movable charge varies inversely as the square of the separation and always points along the line between the charges. We can draw a diagram of the force vectors around the stationary charge, showing the size and the direction of the force on the movable charge at a large number of points. Figure 24 – 4 shows such a set of vectors for both a positive fixed charge and a negative fixed charge. The set of vectors is very much like the set of vectors indicating the gravitational field of the earth (Fig. 14 – 1), except that the gravitational field vectors of the earth all point toward the earth, while these vectors point away from the stationary charge when the force is repulsive and toward it when the force is attractive. The set of force vectors exerted on a movable charge for all points in space is called the force field of the charge; it gives the value of the force felt by the movable charge at any point. Note that the magnitude of each force vector depends on the magnitude of the movable charge, and the direction of the force vector depends on the sign of the movable charge. The drawings in this section show the direction of the force field exerted on a *positive* movable charge. If the movable charge is negative, the force vectors must all point the other way.

Now suppose we have a collection of many charges nailed down at different positions. They exert a force on a movable charge placed at any position, and this force is the vector sum of the forces that each of the nailed-down charges would exert individually (Fig. 24 – 5). The net force on the movable charge in this situation does *not* vary as the inverse square of the separation from any one of the nailed-down charges. But we can compute the net force by adding all the individual forces, or we can measure it experimentally. We can represent it by making a plot of the net force vectors at a large number of positions; and we thus get an idea of the field of force from the whole collection of nailed-down charges.



24–5 (a) Three fixed charges 1, 2, and 3 exert forces on a movable charge. (b) The net force $\vec{F}$ is the vector sum of the forces $\vec{F}_1$, $\vec{F}_2$, and $\vec{F}_3$ exerted by the fixed charges.

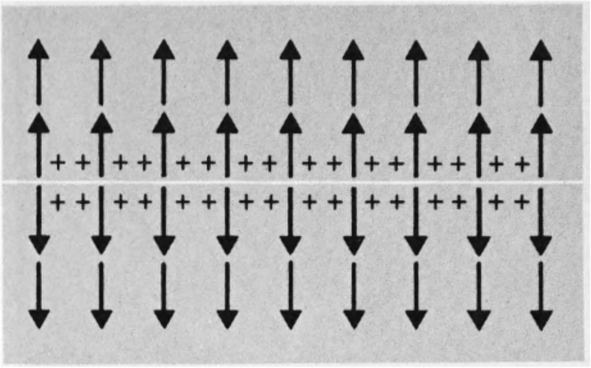


24–6 Two fixed charges A and B on a plane exert forces on a movable charge C. As C moves nearer to the plane, the forces tend to push against each other. So the net force does not increase as much as we might expect. What is the sum of the forces exerted by A and B when C is just in between them?

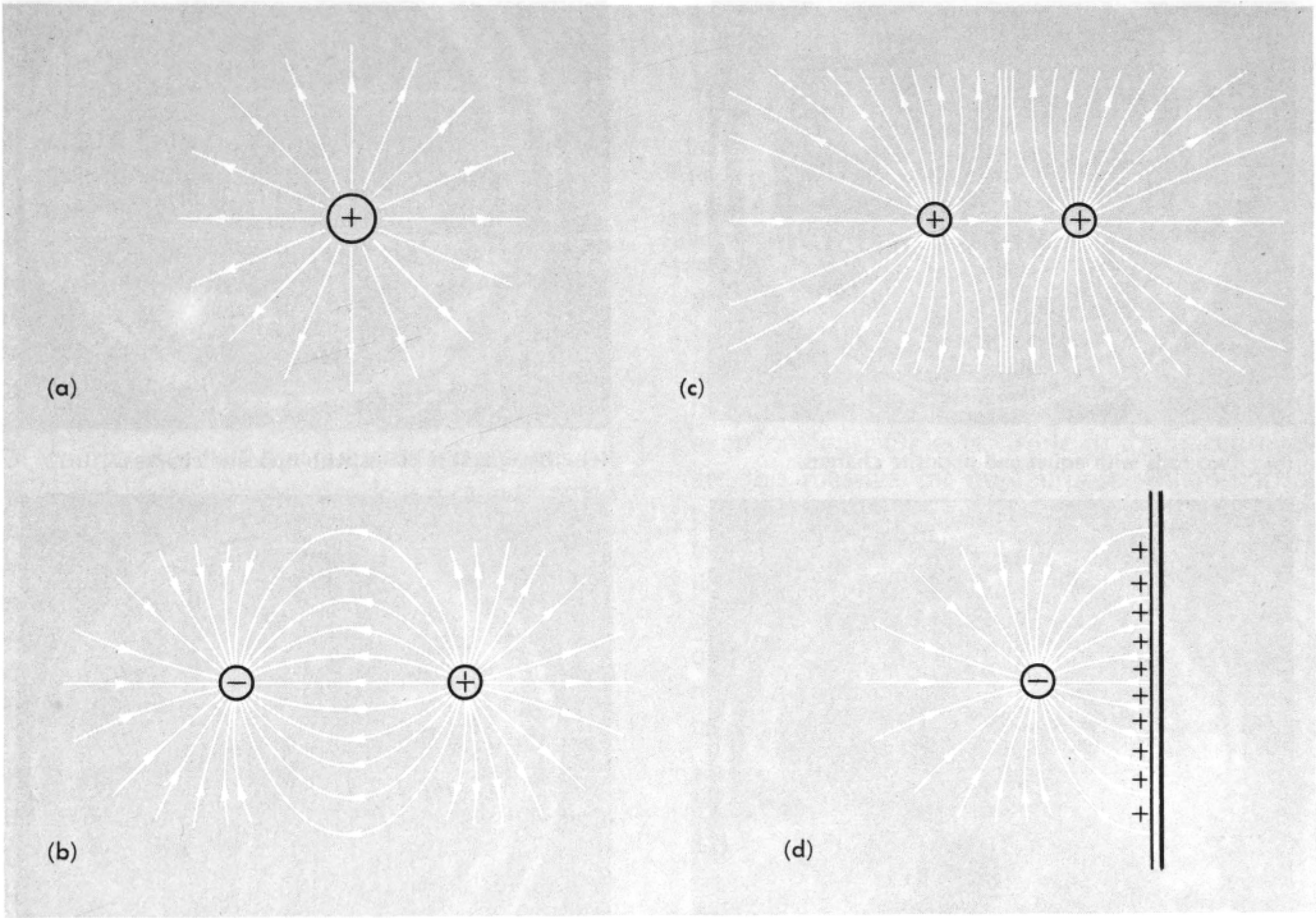
As an example, suppose we have a large number of charges spread evenly on a large plane. We then find that the net force does *not* increase noticeably as we move closer to the plane. Although we are getting nearer to the charges, the forces from them push more and more nearly in opposite directions and become progressively less effective (Fig. 24 – 6). When we actually carry through the vector addition of all the forces, we find that the net force does not change either in magnitude or direction as we approach the plane. The force field is uniform. (A proof is outlined in the box on page 446.) If the charges on the plane and the movable charge are of the same sign, this uniform force field points straight away from the plane. Figure 24 – 7 is a diagram of this field. We shall get an experimental demonstration of its uniformity in the next section.

Often instead of drawing a collection of net-force vectors, we represent an electric force field in a slightly different way by drawing a number of electric field lines. A line is constructed by moving a point continuously in the direction of the force exerted on a movable particle. As the force changes direction, so does the line. The patterns of electric field lines around a single small charged object and around several separated charged bodies are illustrated in Fig. 24 – 8. The direction of each line is the direction of the net force on a movable positive charge placed on the line. Sometimes we can show the electric field lines experimentally by taking advantage of the fact that small oblong bodies will line up along them. For example, we can suspend grass seed in an insulating liquid. Then, if we make an electric field by putting charged objects in the liquid, the seeds line up forming a picture of the field. Figure 24 – 9 shows photographs of the grass-seed patterns of several electric fields. [Compare the center of Fig. 24 – 9 (f) with Fig. 24 – 7.]

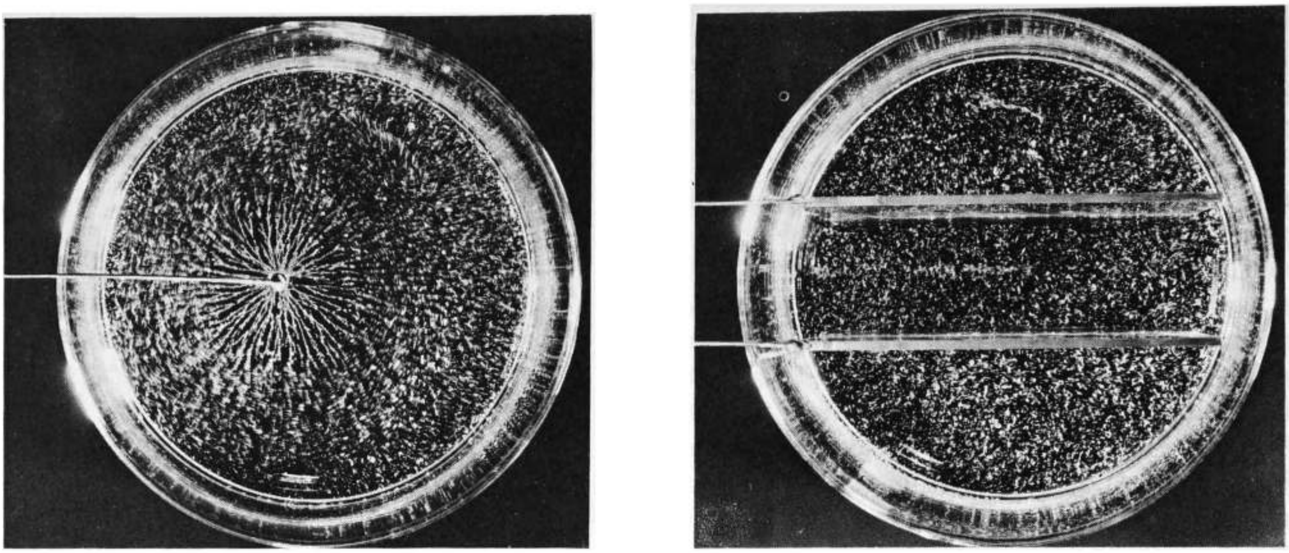
In Fig. 24 – 9 the charged objects are metal conductors. Notice that the electric field lines meet their surfaces at right angles. This is easy to understand. In a metal conductor electrons are free to move from point to point within the metal, but cannot easily leave its surface. If the electric forces which act on the electrons in a metal surface have any component parallel to that surface, they drag the electrons along the surface. This flow of electrons is a current. The currents will continue until the electrons have so arranged themselves that electric field lines become perpendicular to every metal surface. This rearrangement takes a very small fraction of a second; what we measure in these



24–7 The electric force field around an evenly charged plane.

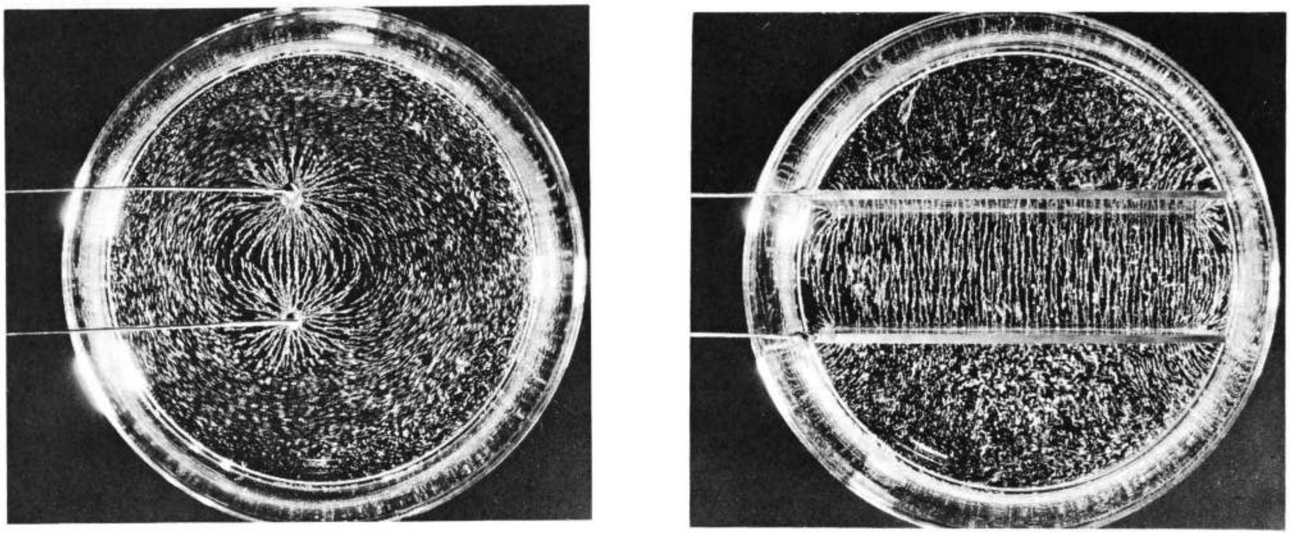


24–8 The electric force field exerted on a positive movable charge by (a) a single positively charged sphere; (b) two charged spheres of opposite sign; (c) two charged spheres of the same sign; (d) a small charged sphere near a large charged plane conductor of the opposite sign. Note that close spacing of the field lines indicates a region where the field is strong. Diverging lines indicate that the field is growing weaker.



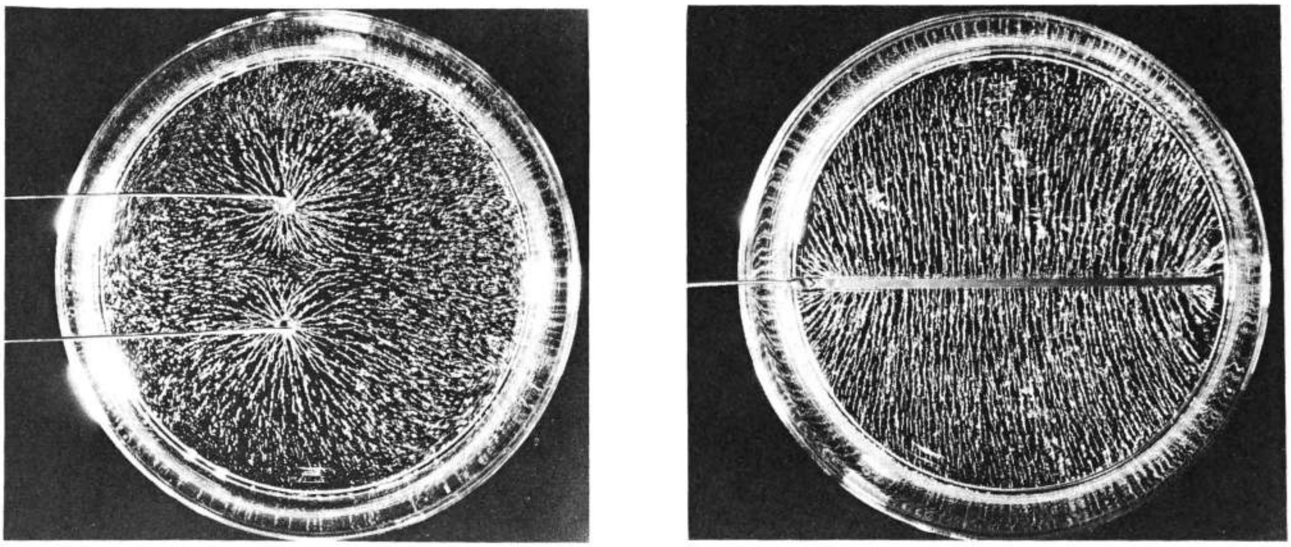
(a) A single charged rod.

(d) Two parallel plates, no electric field.



(b) Two rods with equal and opposite charges.

(e) Two parallel plates with opposite charges.



(c) Two rods with the same charge.

(f) A single charged metal plate.

24–9 Photographs of electric force field patterns formed by grass seeds in an insulating liquid.

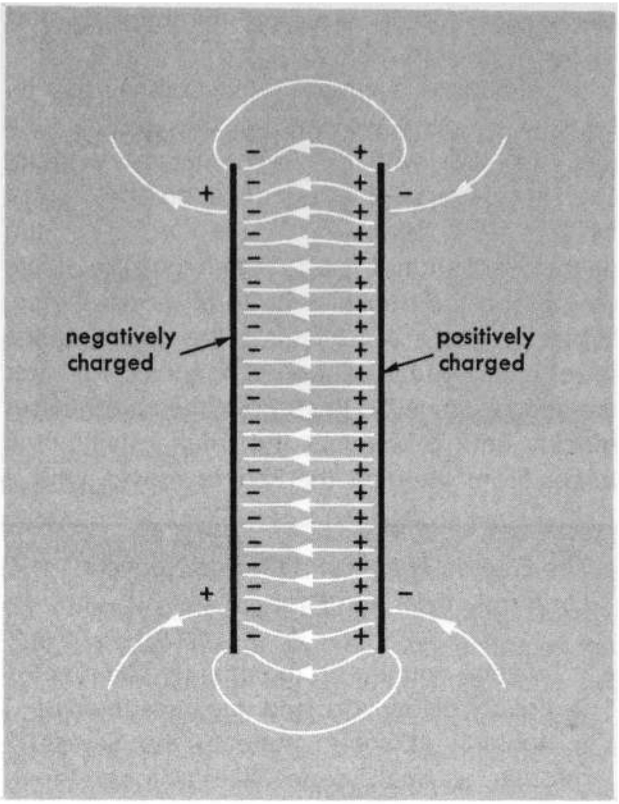
experiments is the field after the charges have stopped moving.

In the next section we are going to use the electric force on a small object between two oppositely charged plane parallel metal plates close together (Fig. 24 – 10). Except at the edges, the charge on each plate spreads out evenly over the inner surface. Consequently, in the region between the plates, the electric field should be uniform and point straight across from one plate to the other. As long as a charge between the plates is not too close to the edges, the force on it should be the same at every point. This force is proportional to the charges on the plates.

24–4 MEASURING SMALL ELECTRIC FORCES

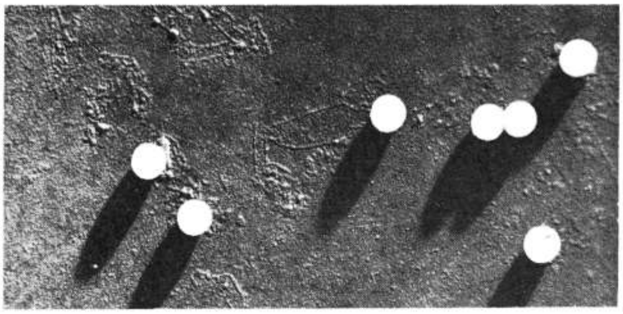
As we shall see, ions have charges of very few particles, often an excess of only one or two. By studying them we shall find that electric particles all have charges of the same size. Charge comes in separate grains; and every grain, whether positive or negative, is of the same magnitude as any other. The identity of the grains of charge, like the identity of atoms of matter, provides us with natural units. Their discovery is one of the great landmarks distinguishing modern physics from ancient. It allows us to specify the charge on an object by counting the extra number of electric particles of one sign as compared with those of the other sign. It makes our picture of charge simple. Of course, the fact that elementary charges all have the same magnitude does not guarantee that particles are the same in all other respects. As we shall see later, there are particles of several different masses, all with the same charge.

For the moment we shall study the electric forces on ions, and then use them to establish the units of charge. The electric forces we can exert on ions are small; these forces will not move ordinary-sized objects or stretch springs appreciably. In any case, we cannot tie single ions to a spring or be sure that we have only one ion on any object of visible size. We must therefore find a way of handling really tiny objects with an excess of only a few elementary electric particles on them. In this section, we shall discuss one method of manipulating such objects and of measuring the minute electric forces we can exert on them. Then we shall be equipped to understand the evidence given in the next section for the natural unit of charge.



24–10 The force field exerted on a positive charge by two oppositely charged metal plates placed together, and parallel to each other. The force on the charge is the same anywhere between the plates, except near the edges.

We need objects which are much larger than ions or gas molecules and yet small enough to be moved about by the electric forces we can exert on ions. Objects visible to the naked eye are too massive for our purpose. Microscopic plastic spheres about 10^{-6} meter in diameter turn out to be about right (Fig. 24 – 11). Although they are large enough to be seen as spots of light through an optical microscope, the mass of each is only about 10^{-15} kg, or a micro-microgram. Since these spheres are manufactured for calibrating the distances seen with



24–11 An electron-microscope photograph of a few plastic spheres magnified about 2500X.

electron microscopes, a large number of almost identical spheres are available.

In air these spheres move rather slowly. Looking at them, we see that they fall steadily (Fig. 24 – 13) with no apparent acceleration. On the average, the net force on them must be zero. This zero net force is the vector sum of the gravitational force, which pulls a sphere down, and the opposing force of air resistance. When a sphere starts falling, the air resistance is zero, but the faster the sphere goes the greater the air resistance. So the drag of the air increases quickly until it balances the small gravitational force. From then on the sphere moves with a constant velocity, called the terminal velocity.

If we place some of these plastic spheres between two charged plates, we find that there is an electric force acting on the spheres, in addition to the gravitational force. Some of the spheres move upward, others fall downward. Apparently, the spheres carry an electric charge. As a sphere starts to move under this net driving force (the vector sum of the gravitational and electric), the air resistance will change until the sphere will again move with constant velocity. This terminal velocity will be different from the terminal velocity with only the gravitational force acting.

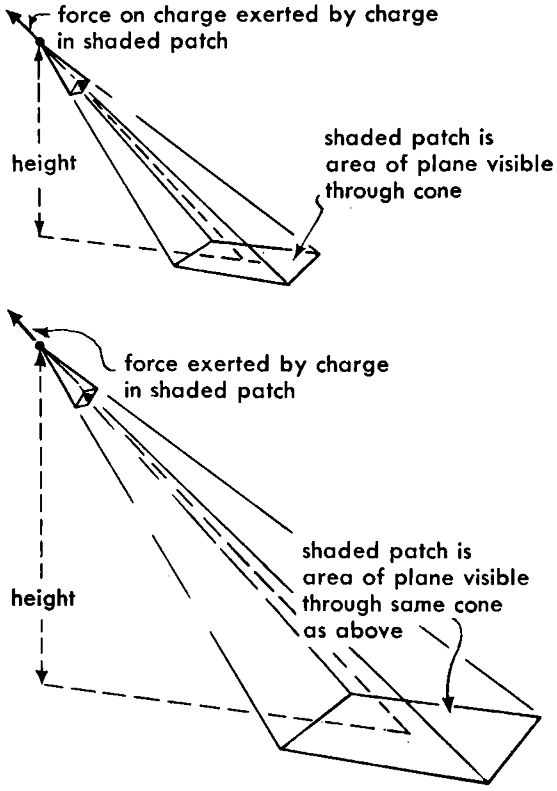
The Electric Force on a Charge Above a Uniformly Charged Plane

Imagine you are on the charge. The plane is directly below you; and you look down on it, looking at some angle to the vertical through a small cone—like a long, thin megaphone (Fig. 24 – 12). The area of the little patch of plane you see through the cone increases with the square of your height above the plane. For example, if you move twice as far from the plane, you see a patch of twice the original length and twice the original width on the plane. The patch has four times the original area, and therefore the charge of the patch is four times the original charge seen at that angle. Also, this charge is twice the original distance away. So the force from a given amount of charge is one quarter what it was originally. The ratio of the force at a given angle to the original force at that angle is changed by two factors, the factor 4 from the fourfold charge and the factor $\frac{1}{4}$ from the square of the doubled distance. The net effect—because $4 \times \frac{1}{4} = 1$—is to give exactly the original force.

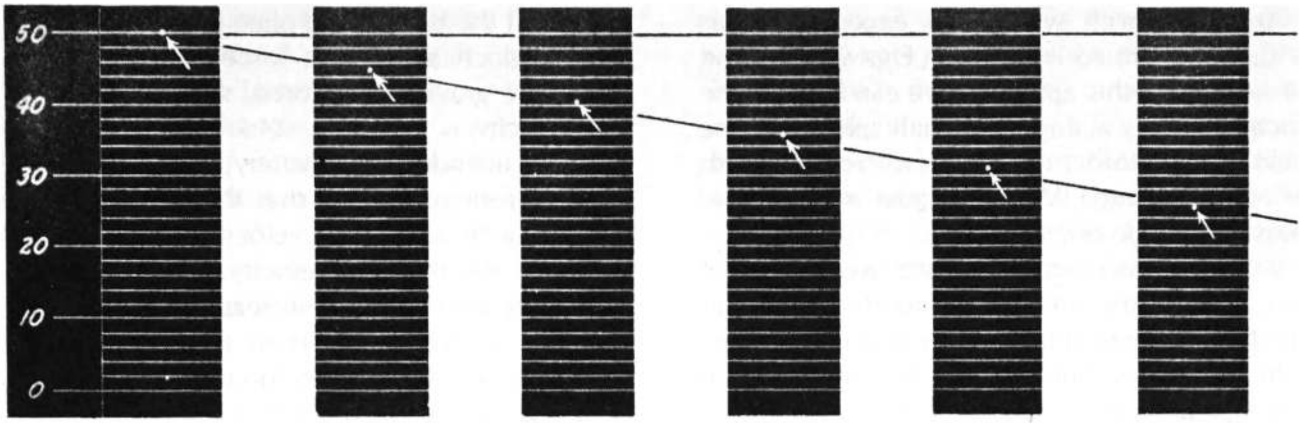
The falling off of the force from a given amount of charge and the increase in the amount of charge seen through the megaphone always compensate — $1/r^2$ and r^2. Therefore the force coming up at any angle is the same at all heights as long as we do not have to look off the edge of the plane.

Finally, since all the forces coming up at all angles stay the same, the net force also stays the same. It does not depend on the height above the plane. Note that with any other force law you would not obtain a uniform force field. The $1/r^2$ force law is the only one that will cancel with the geometrical r^2. Consequently, the observation of a uniform force field in this case verifies Coulomb's law.

Of course, at really great heights, so great that the whole charged plane looks like a point, the force falls off as $1/r^2$. The force starts to decrease appreciably when we get about $\frac{1}{20}$ as far away as the smallest dimension of the charged region.

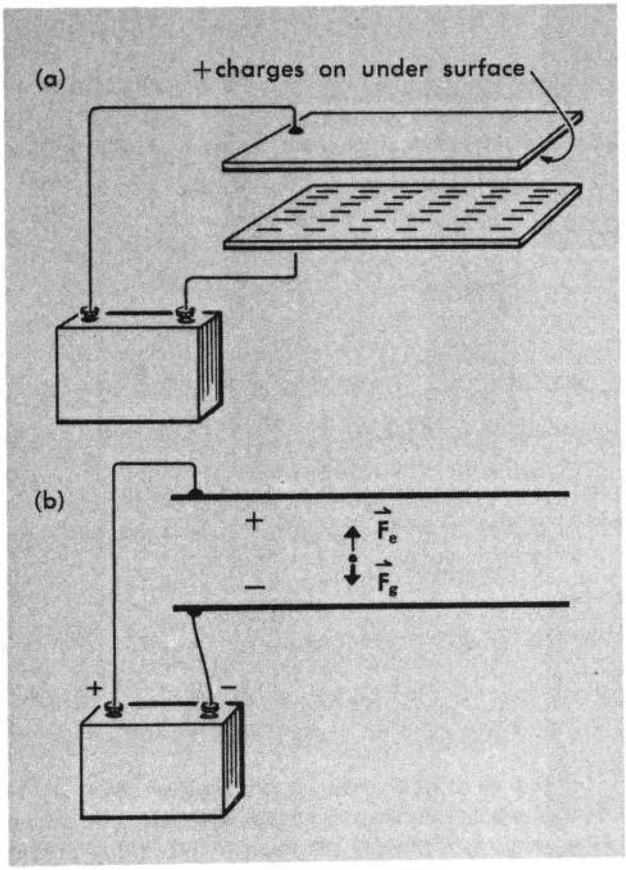

force on charge exerted by charge in shaded patch
height
shaded patch is area of plane visible through cone
force exerted by charge in shaded patch
height
shaded patch is area of plane visible through same cone as above

24–12 Looking down at a plane through a narrow cone, you see an area of the plane that depends upon your height above it.

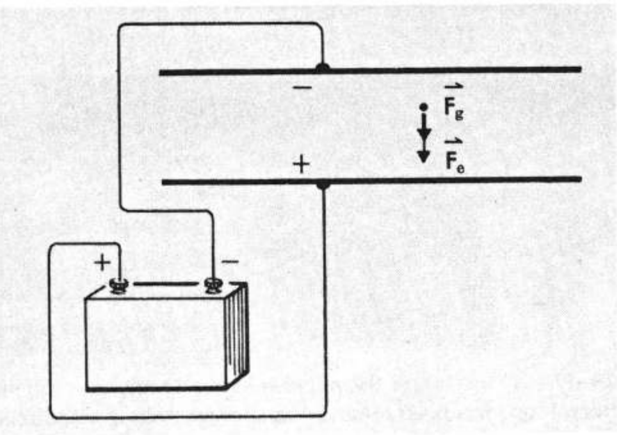


24–13 Photographs of a plastic sphere driven downward by gravity, taken through a microscope at one-second intervals. The sphere falls about the same distance in each interval across a grid slightly over 1 mm long.

How is the terminal velocity related to this net driving force? Let us consider a sphere placed between two charged metal plates connected to a battery, as shown in Fig. 24 – 14. If we vary the electric charge on the plates, we can change the electric force acting on the sphere. By choosing the right battery and plate separation, we can have an electric force acting on the sphere which will be equal in magnitude but opposite in direction to the gravitational force. Thus, the sphere will be held in balance between the plates. What will happen if we reverse the connections of the two plates with the terminals of the battery (Fig. 24 – 15)? The electric force is still equal in magnitude to the gravitational force, but now it points in the same direction as the gravitational force. Thus, the net driving force on the sphere is now twice the gravitational force alone. If we measure the terminal velocity with which the sphere moves, we find it is twice the terminal velocity with which the sphere moved under gravitation alone. This result suggests that the terminal velocity of the sphere is directly proportional to the driving force. This relation has been checked and holds over a wide range of velocities.



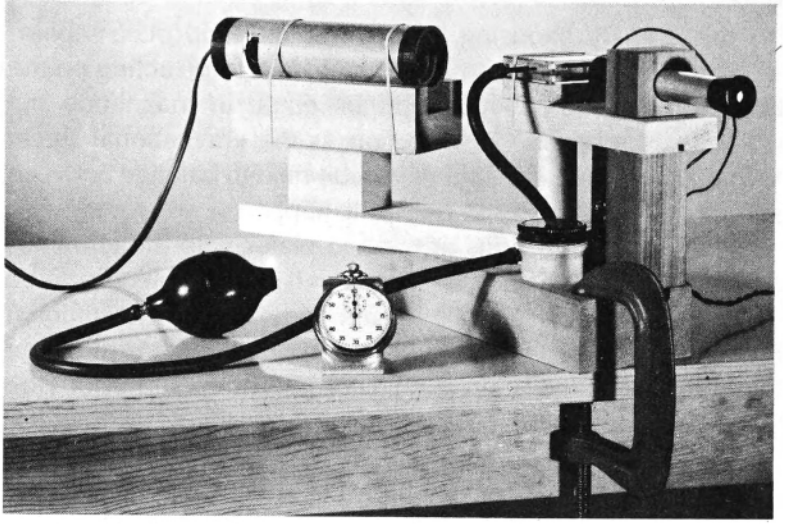
24–14 By choosing the right battery and the right plate separation, we can hold a sphere motionless between the plates. The electric force and the gravitational force must be equal: (a) shows the way the plates and the battery are arranged; (b) is a sketch of the same apparatus with schematic plates.



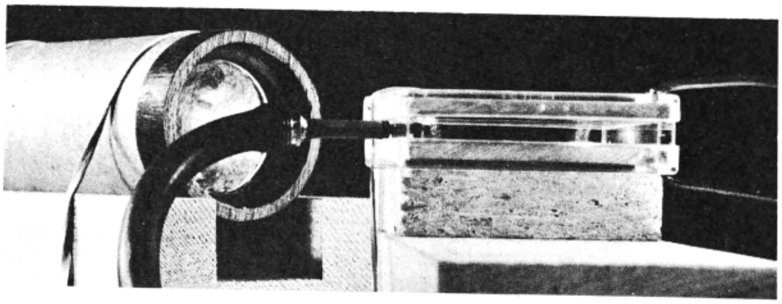
24–15 With the connections of the balancing battery reversed, the electric force now acts downward. The driving force is twice the gravitational force.

Apparatus with which this experiment was actually performed is shown in Figs. 24 – 16 and 24 – 17. With this apparatus we can balance the force of gravity acting on a small sphere, whose mass is of the order of a micro-microgram, with an electrical force. We shall refer to the apparatus as a micro-microbalance.

With the micro-microbalance we can show that the electric force is proportional to the number of identical batteries connected in series with the plates. Suppose we first measure the terminal velocity of a plastic sphere when no electric force acts on it [Fig. 24 – 18 (a)]. Then we select the battery and plate separation which will produce an electric force which just balances the gravitational force, so that the terminal velocity is zero [Fig. 24 – 18 (b)]. We now connect an additional battery, identical to the first, in series and find that the sphere moves upward with a terminal velocity equal and opposite to the terminal velocity with no electric force [Fig. 24 – 18 (c)]. This leads us to conclude that the electric force when two batteries are used is twice the electric force when only one is used. Indeed, if we add a third battery in series to the other two, the sphere moves up-



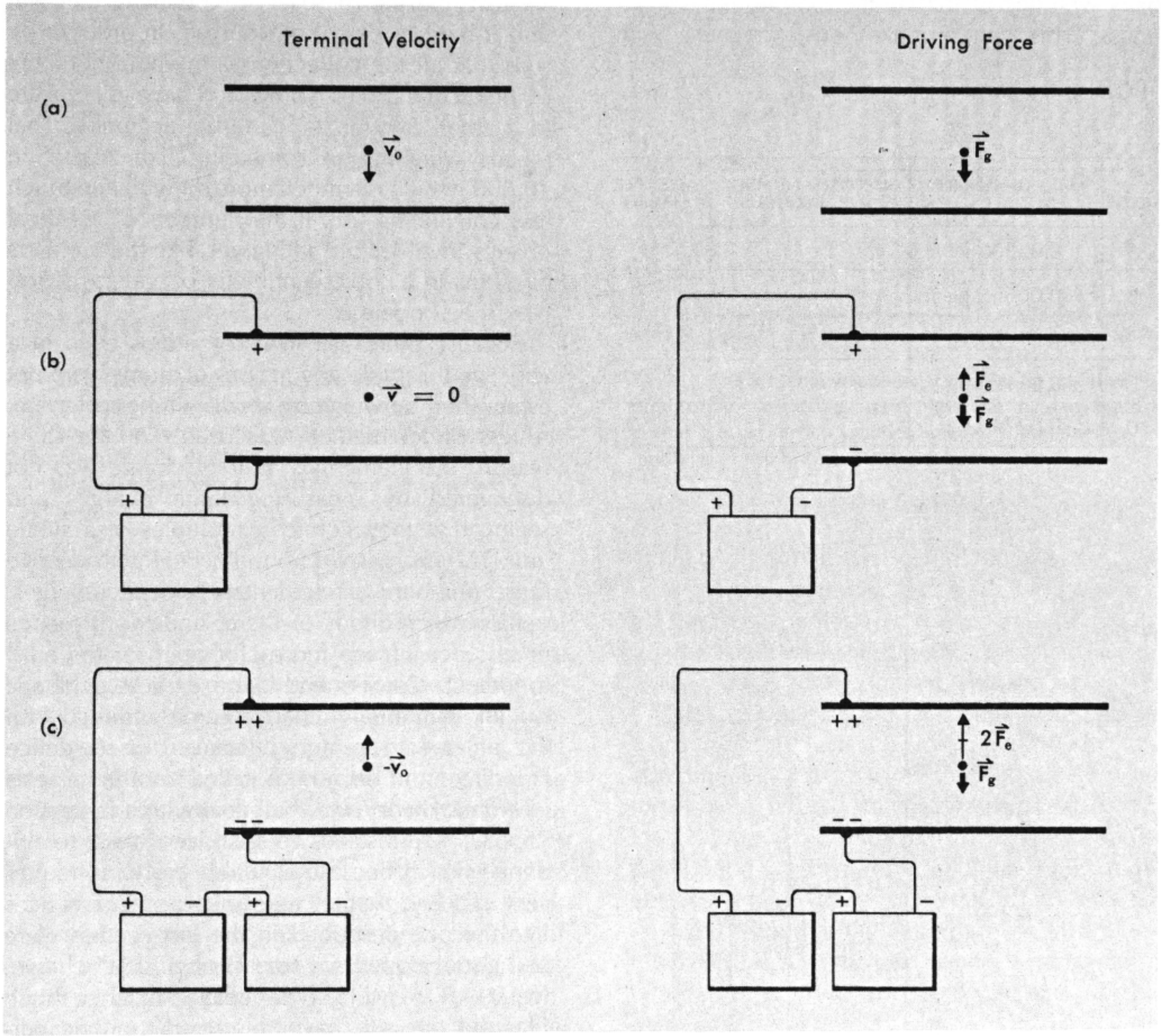
24–16 The micro-microbalance. The large bulb is squeezed to force plastic spheres out of the jar, up through the tube, and into the space between the two metal plates, which are connected to batteries by the wires at the right. A bright light source, at the left center of the picture, illuminates the spheres so that they can be seen through the microscope at the right.



24–17 A close-up of the micro-microbalance plates. The microscope has been removed to show the plates and their connections to the batteries more clearly.

ward with a terminal velocity of twice the magnitude of the terminal velocity with no electric force. While one battery just cancels the force of gravity, the other two produce a net force on the sphere which is double in magnitude and opposite in direction to that of the gravitational force.

Let us note one more thing. As we haul the charged sphere up and down between the metal plates, it moves at the same terminal velocity, no matter where it is between the plates. The electric force is therefore the same at all points between the plates. This is an experimental proof of the constancy of the electric force field between charged plates. It agrees with the theoretical conclusion stated in the last section where we asserted that the field *should* be constant.



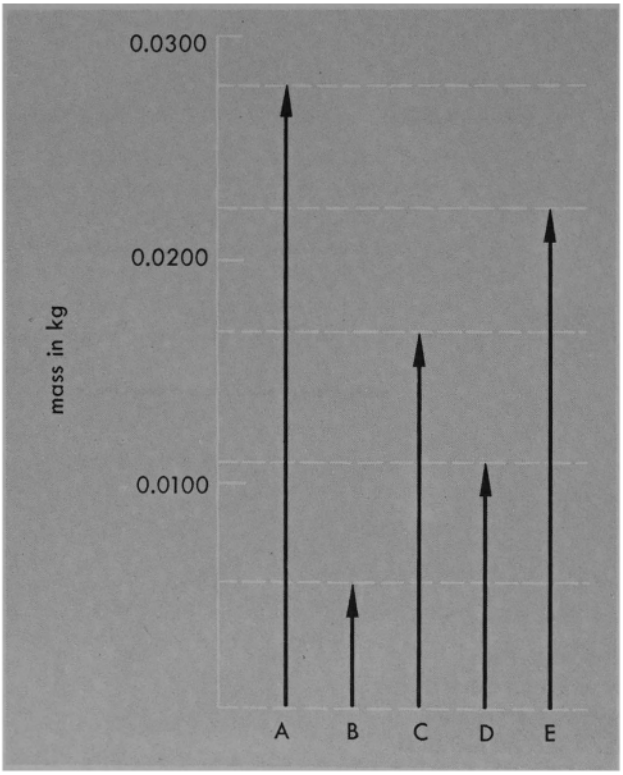
24–18 We can show that the electric force on a sphere is proportional to the number of batteries in series by (a) measuring the terminal velocity when no electric force is acting, (b) balancing the gravitational force on the sphere with an equal and opposite electric force, and (c) measuring the terminal velocity after a second identical battery has been added in series. The final terminal velocity is equal and opposite to the terminal velocity in (a), showing that now the electric force on the sphere must have twice the magnitude of the gravitational force.

24–5 THE ELEMENTARY CHARGE

Suppose somebody gives you a set of stiff paper bags, sealed shut, with various numbers of marbles in them. He challenges you to tell whether the marbles he put in are all alike and how many are in each bag—without breaking the seals. You weigh the bags and find the masses for the marbles given in Table 1. By plotting these masses as in Fig. 24 – 19, you see that they are all multiples of 0.00560 kg. Then you might very well decide that all the marbles have equal mass. This would make sense, because with marbles of 0.00560 kg there would be 5 in bag *A*, 1 in bag *B*, 3 in bag *C*, 2 in *D*, and 4 in *E*. (You would probably rattle bag *B* to see whether your idea was correct.)

TABLE 1

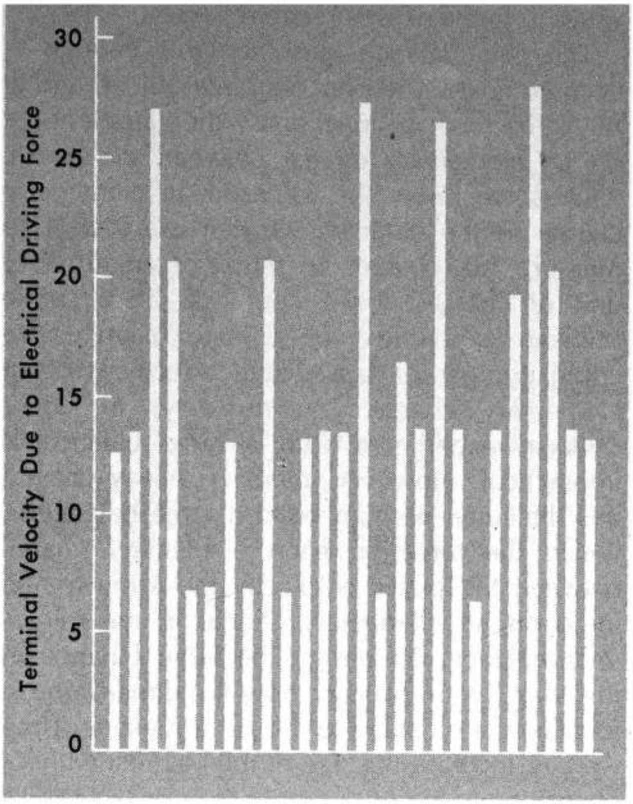
BAG	MASS OF MARBLES IN THE BAG	SUPPOSED NUMBER OF MARBLES	MASS PER MARBLE
A	0.02800 kg		
B	0.00560 kg		
C	0.01680 kg		
D	0.01120 kg		
E	0.02240 kg		



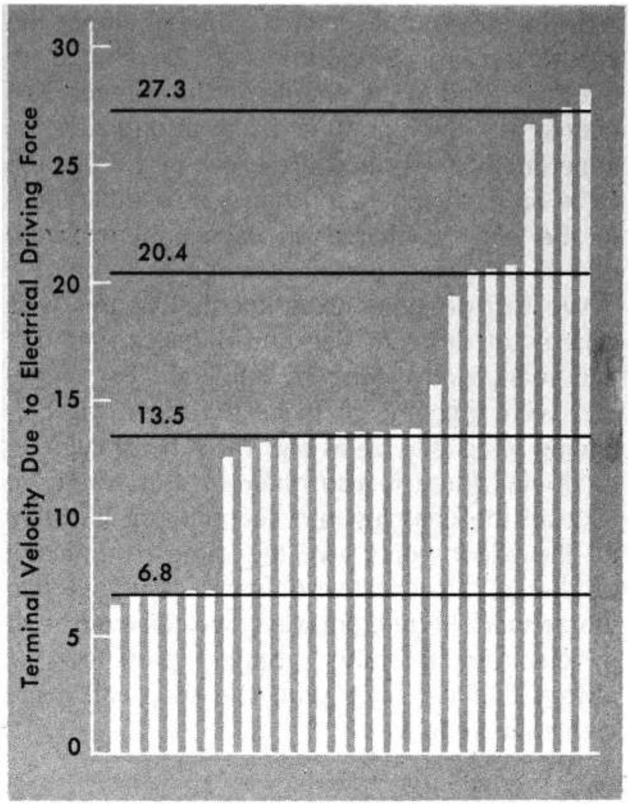
24–19 The masses of the five bags of marbles. The horizontal broken lines are to make comparison easier.

On the other hand, if there are huge numbers of marbles in each bag, you probably will not suspect that every marble has the same mass. For example, suppose that the collections of marbles have masses of 5.76 kg, 6.27 kg, 3.24 kg, 4.82 kg, etc., that is, about 1030, 1120, 580, 860 . . . marbles. With such masses, even if you suspect that the marbles are identical, you will find it hard to convince yourself. In order to be sure that all the collections are whole numbers of identical marbles, you would have to measure to a small fraction of 1-marble accuracy. That means weighing to something like 1 part in 10,000, which is a much more difficult and much less convincing job. If the numbers of identical objects involved are still larger, like the numbers of atoms in a cubic centimeter of water, the job would be hopeless.

Because there are so many atoms even in a very small sample, the masses of atoms were not established by weighing small numbers of them. A less direct method was used. For the same reason, the elementary unit of charge was not discovered by measuring large charges and seeing that they come in multiples of a single unit. The charges we normally deal with involve huge numbers of elementary electric charges, millions of millions of them, and the presence or absence of one more charge is far too hard to detect. Consequently, the earliest evidence for an elementary charge came indirectly, in the nineteenth century. Because this evidence is indirect and because it is tied to other aspects of atomic theory, we shall deal with it in Section 25 – 3. Here instead we shall look at some evidence which became available early in the present century. At that time micro-microbalances like the one described in the last section were being developed for measuring ionic charge. In 1909 R. A. Millikan perfected a balance much like the one we have described, and he performed a series of experiments which clearly show the existence of the elementary unit of charge. Where we use plastic spheres of known mass, he used tiny oil drops—and he had to measure their masses while examining their charges. Our experiment is therefore easier than



24–20 The terminal velocities (measured in units convenient for this experiment) due to the electric part of the driving force acting on several plastic spheres. The terminal velocity due to the gravitational part of the driving force has been subtracted from the observed values.



24–21 The terminal velocities of Fig. 24–20 replotted in order of increasing magnitude. Note that the magnitudes cluster around certain values which are whole-number multiples of the lowest one.

his, but otherwise it is nearly the same.*

Suppose that we place several spheres like those shown in Fig. 24 – 11 between the plates of the micro-microbalance. When the plates are uncharged and no electric force is acting, all the spheres have the same terminal velocity. This occurs because the spheres have the same mass, so the driving force due to gravity is the same on each sphere. On the other hand, when the plates are charged, the terminal velocities are no longer the same for each sphere. The driving force is now part electric and part gravitational, so the terminal velocity is the vector sum of a part due to the electric and a part due to the gravitational driving force. If we subtract the terminal velocity due to the gravitational driving force, we will find the terminal velocity each sphere would have if the only driving force were electric. Since we know that the terminal velocity is proportional to the driving force, we will be able to determine the electric force acting on each sphere.

The terminal velocities of several spheres were measured in an experiment using the equipment of Figs. 24 – 16 and 24 – 17. The terminal velocity due to the gravitational driving force alone was subtracted from each observed value, and the resulting velocities due to the electric driving force alone are plotted in Fig. 24 – 20. Note that the values are not distributed randomly. They cluster around certain values which are multiples of the lowest one. For this particular experiment, the velocities cluster around 6.8, 13.5, 20.4, and 27.3 (in units which were convenient to use in this experiment), as seen by rearranging them in order of magnitude (Fig. 24 – 21). Since the terminal velocity is proportional to the driving force, we conclude from these data that the electric force is also distributed around certain values. Now, the electric force is proportional to the charge on the plates and to the charge on the spheres. The charge on the plates is fixed throughout the experiment; thus, the differences in the electric force on each sphere must arise from differences in the electric charge on each sphere. These differences come in multiples of a basic unit.

---

*In Millikan's time there were no standardized plastic spheres and no electron microscopes with which to see that they are all the same. He was forced to do something different.

We can do other experiments like this with spheres of different size and different material, and many were done by Millikan. For example, with a sphere of twice the mass, the balancing force is twice as big. The ratio of velocity to driving force is also different. But when we take all the changes into account, we find that all the experiments give the same natural unit of charge. We can use different gases; we can ionize them in different ways; we can even send in electrons or other charged particles—we always find the same unit. We shall call it the elementary charge.

24–6 A LARGE ELECTRIC BALANCE

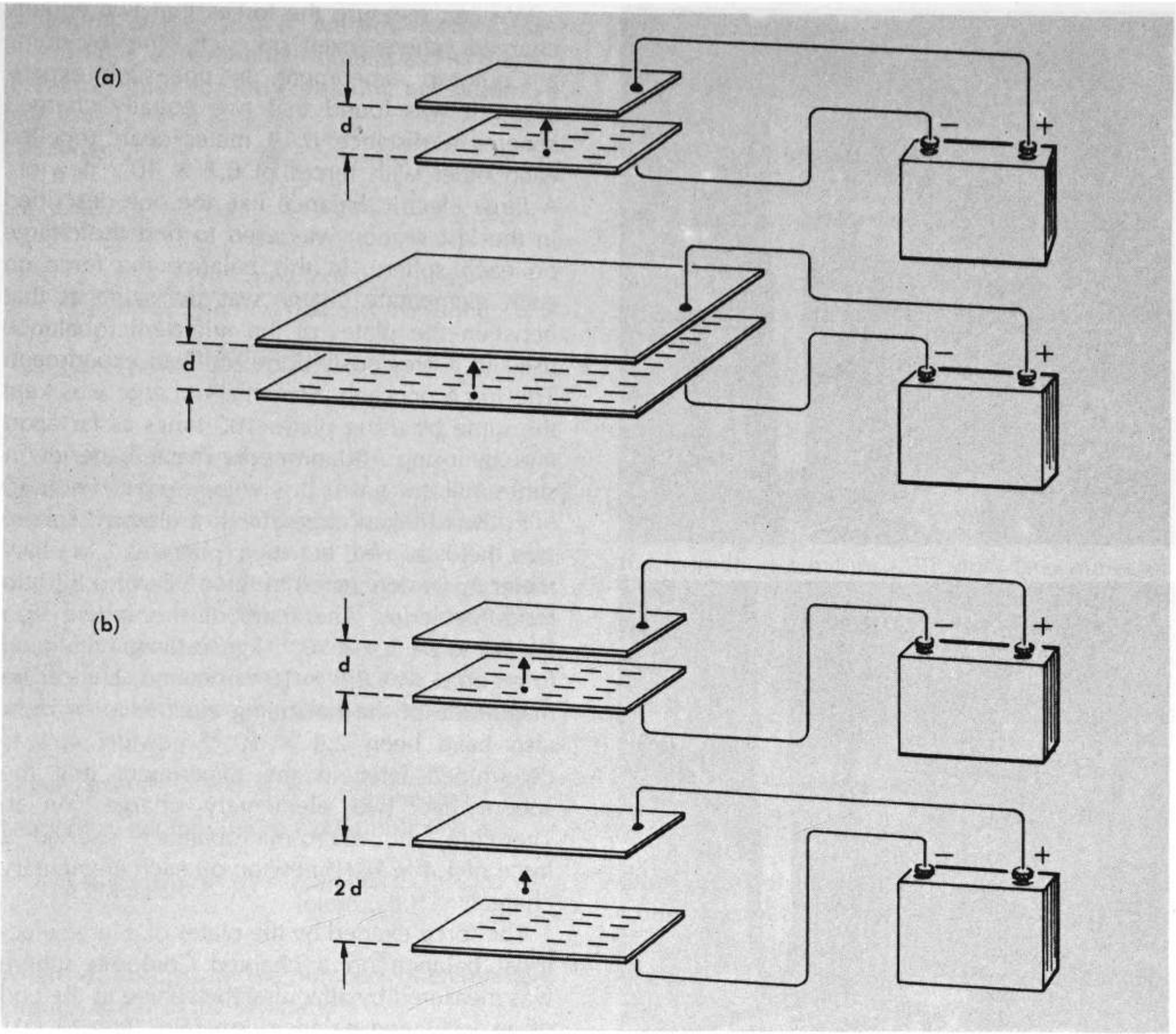
With a micro-microbalance we can measure the charge of an object as a certain number of elementary charges. We know that this method will work when the number of elementary charges is small and the object is microscopic. Will it also work for large numbers of elementary charges? A really large number of elementary charges will not stay on a microscopic object. On the other hand, if the object is large enough to retain a large amount of charge, it will be too large for the micro-microbalance. Consequently, the micro-microbalance will not work for large numbers of elementary charges.

How, then, are we going to measure a large charge as a number of elementary charges? There are several ways, but the most obvious is to build a bigger balance. We want a balance large enough to accommodate a sphere from a Coulomb experiment (see Sections 24 – 1 and 24 – 2) or some similar object. If we can measure the charge on such an object in terms of the number of elementary charges, we should then be able to measure all charges as numbers of elementary charges. We shall see in the next section that once we have measured one big charge, the Coulomb experiment will let us measure any other in terms of elementary charges.

The new balance must be large enough to accommodate a reasonably large object, and at the same time we must know the electric force on an elementary charge between the plates. (Unless we know the force on an elementary charge in the balance, we can use it only to measure big charges in terms of an arbitrary unit of charge.) If we could do an accurate Millikan experiment in a large balance, we could compare a big charge directly with an elementary charge. Unfortunately, to do a Millikan experiment with a large balance is impractical. However, there is a way around this difficulty: we can build a large balance in which the force on each elementary charge between the plates is exactly the same as the force on an elementary charge in the micro-microbalance. To design the large balance in this way, we must learn how the dimensions of a balance are related to the force it exerts. Then we can increase the size of the balance without changing the force.

In order to scale up a balance, we shall have to change both the area of the plates and their separation. We can learn the effect of each of these changes by experiment, both with the micro-microbalance and with widely separated charged plates exerting forces on large charges.

When we use plates of larger area at the same distance apart and connected to the same battery (Fig. 24 – 22), the force on a charge between them is the same. However, if we again keep the same battery connected, but now increase the separation between the plates, the force decreases, going in inverse proportion to the separation. We can compensate for this by using more batteries. Since the force increases in proportion to the number of batteries (Section 24 – 4), we must double the number of batteries whenever we double the separation of the plates. Direct tests demonstrate the same thing. As long as the plates are big compared with their distance apart, we keep the same electric force by increasing the separation and increasing the number of batteries in the same proportion. The exact area of the plates does not matter. Therefore, to scale up the electric micro-microbalance to a convenient size, we use big parallel metal plates a convenient distance apart. And we make sure that the elec-



24–22 **(a) If we increase the area of the plates without changing the separation or the battery, the force on an elementary charge anywhere between the plates remains the same. (b) If we double the distance between the plates with the same battery connected, the force is halved. In general, the force is proportional to 1/d, where d is the distance between plates.**

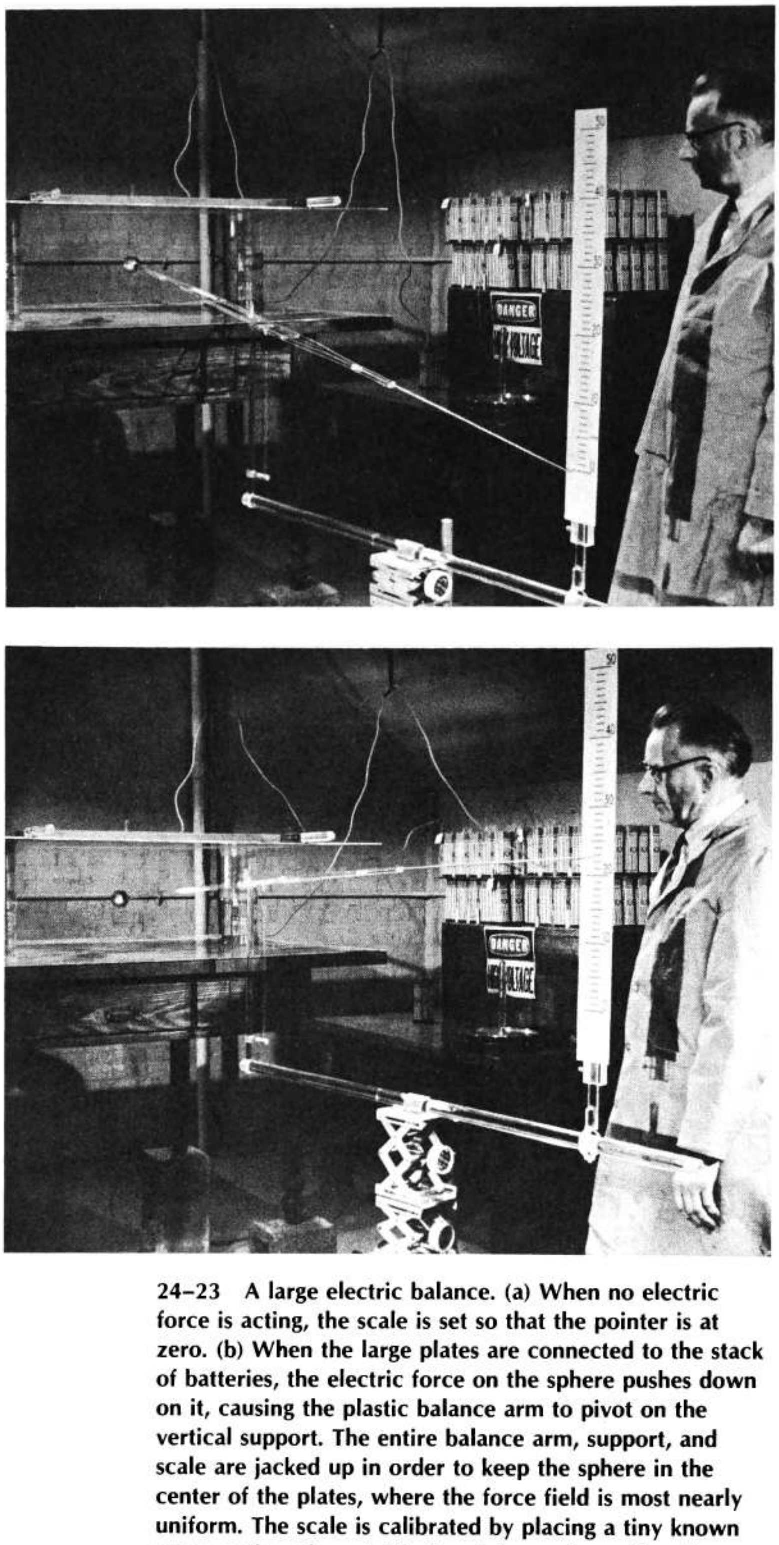
tric force on an elementary charge between the plates is the same as in the micro-microbalance by charging the plates with a sufficient number of batteries in series.

24–7 THE CONSTANT IN COULOMB'S LAW

When we discussed Coulomb's law and the electric charge in Sections 24 – 1 and 24 – 2, we found that

$$F = k\,\frac{q_1q_2}{r^2}$$

where F is the force one charge exerts on the other, q_1 and q_2 are the charges, and r is the separation between them. The proportionality factor k is a constant. Its value depends on the units in which we measure force, separation, and charge. Now we are in a position to evaluate the constant k in Coulomb's law.



24–23 A large electric balance. (a) When no electric force is acting, the scale is set so that the pointer is at zero. (b) When the large plates are connected to the stack of batteries, the electric force on the sphere pushes down on it, causing the plastic balance arm to pivot on the vertical support. The entire balance arm, support, and scale are jacked up in order to keep the sphere in the center of the plates, where the force field is most nearly uniform. The scale is calibrated by placing a tiny known mass on the sphere (with the plates uncharged) and noting how many divisions correspond to the known gravitational force on the mass.

We can measure the forces that two equally charged spheres exert on each other by doing a Coulomb experiment. In one such experiment, it was found that two equally charged spheres a distance 0.15 meter apart repelled each other with forces of 6.7×10^{-4} newton. A large electric balance like the one described in the last section was used to find the charge on each sphere. In this balance the force on each elementary charge was the same as that between the plates of the micro-microbalance used in a previously done Millikan experiment. The force on each elementary charge was kept the same by using plates 100 times as far apart and by using 100 times as many batteries in series.

In the Millikan experiment, a charged sphere was held at rest between plates 3.1×10^{-3} meter apart connected to three 90-volt "B" batteries in series. The mass of the sphere was known to be 2.9×10^{-15} kg, so the gravitational force on it was 2.8×10^{-14} newton. Hence the magnitude of the balancing electric force must also have been 2.8×10^{-14} newton. It was determined later in the experiment that the sphere had two elementary charges on it. Therefore, the micro-microbalance exerted a force of 1.4×10^{-14} newton on each elementary charge.

The force exerted by the plates of a large electrical balance on a charged Coulomb sphere was measured by attaching the sphere to the end of an insulating balance arm [Fig. 24 – 23 (a)]. First, with the plates uncharged and not connected to any batteries, the position of the scale was adjusted so that the pointer on the balance arm was at zero. Then, with the plates connected to three hundred 90-volt batteries, the change in the pointer's position due to the electric force on the sphere was noted [Fig. 24 – 23 (b)] and was found to correspond to a force of 3.55×10^{-3} newton. Since the plates exerted a force of 1.4×10^{-14} newton on each elementary charge, the charge on the sphere must have been

$$q = \frac{3.55 \times 10^{-3}\ \text{newton}}{1.4 \times 10^{-14}\ \text{newton/elem. ch.}}$$

$$= 2.5 \times 10^{11}\ \text{elementary charges}$$

In the Coulomb experiment described at the beginning of this section, we now have values for all the experimental quantities. We can, therefore, solve for the constant k in Coulomb's law.

We have

$$k = \frac{Fr^2}{q_1 q_2}$$

$$= \frac{(6.7 \times 10^{-4}\ \text{newton})(0.15\ \text{meter})^2}{(2.5 \times 10^{11}\ \text{elem. ch.})(2.5 \times 10^{11}\ \text{elem. ch.})}$$

$$= 2.4 \times 10^{-28}\ \frac{\text{newton-m}^2}{(\text{elem. ch.})^2}$$

We can check on the value of k by doing other Coulomb experiments and using the large balance to measure the charges. As we expect, the value of k always comes out the same. It is one of the universal constants of nature. A great deal of work with all sorts of experiments has gone into measuring k accurately. The results of all these experiments are in close agreement, and the value given by the best experiments is almost exactly $2.306 \times 10^{-28}\ \frac{\text{newton-m}^2}{(\text{elem. ch.})^2}$. Consequently, we now write Coulomb's law as

$$F = \left(2.306 \times 10^{-28}\ \frac{\text{newton-m}^2}{(\text{elem. ch.})^2}\right) \frac{q_1 q_2}{r^2}$$

where q_1 and q_2 are the numbers of elementary charges and r is the separation.

As we remarked in the last section, once we can measure an ordinary charge by counting the elementary charges, we should be able to devise a procedure for measuring any charge in terms of elementary charges without going to the extreme of bringing each charge into a large electrical balance. Our knowledge of the value of k is the key to measuring charges. By charge-sharing we can always get two equal charges. After measuring the force between them at a definite distance we know F, r, and k in Coulomb's law, and we can therefore work out q in elementary charges. Using one of the charged bodies as a standard, we can then measure the number of elementary charges on any other body by measuring the force between it and the standard.

In addition to giving us a practical way of counting the numbers of elementary charges by measuring electric forces between ordinary-sized objects, the constant $2.306 \times 10^{-28}\ \frac{\text{newton-m}^2}{(\text{elem. ch.})^2}$ tells us directly the force between two elementary charges one meter apart. They exert just 2.306×10^{-28} newton on each other. No doubt that force looks absolutely negligible. But in fact it is huge. The gravitational attraction between two hydrogen atoms at this distance is about 2×10^{-64} newton. The electric force between two elementary charges is 10^{36} times as big. (If you say this in terms of millions of millions, you have to say the word "million" six times over.) The force is so big that two moles of elementary charge (two collections of 6×10^{23} charges) placed one on each side of the earth would push on each other with a force of half a million newtons—50 tons. Two moles of the heaviest atoms placed at opposite ends of a diameter of the earth would have no appreciable gravitational effect on each other.

24–8 CONSERVATION OF CHARGE

Our picture of particles of electric charge leads to the idea of the conservation of charge. When we charge bodies by rubbing them, we transfer charged particles from one to another. What one body gains in charge the other one loses. When we charge the plates of an electric balance with a battery, the battery transfers charge from one plate to the other. The charges of the plates are equal and opposite.

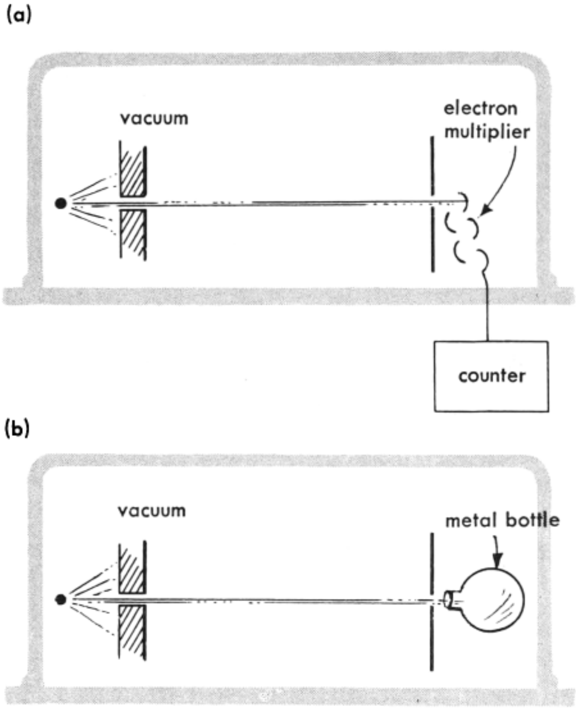
In certain very unusual circumstances we can "create" charged particles, but we shall find that we always create them in pairs, and the charge of one particle is equal and opposite to the charge of the other. Sometimes nature "creates" charged particles automatically: for instance, a neutron turns into a proton and an electron. The proton and electron thus created have equal and opposite charges. The total charge is zero before and after the creation.

All our evidence indicates that the total amount of charge never changes. Like the conservation of energy, the conservation of charge is a law of nature which extends over everything we know.

24–9 THE ELECTRIC CHARGE OF ELECTRONS AND OTHER PARTICLES OF MATTER

We already know that electrons have a negative electric charge, but we have yet to find out whether their charge equals one or several elementary charges. We can determine the charge of an electron by counting electrons as we add them to a neutral body and then measuring the charge on the body in numbers of elementary charges.

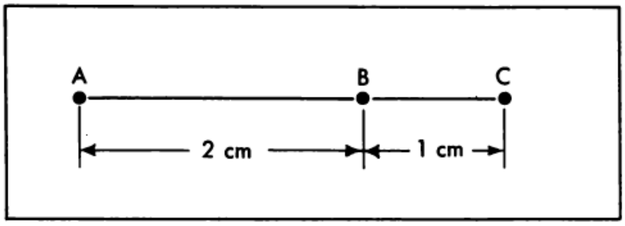
With an electron gun (Section 23 – 13) we can produce a beam of high-speed electrons which will make a luminous spot on a fluorescent screen. If the number of electrons passing per second in the beam is small enough, the individual electrons will cause individual scintillations, pinpoint flashes of light. Although each scintillation gives off very little light, the scintillations can be detected by looking at them with a photoelectric cell. Thus the electrons can be counted individually. The electrons can also be counted directly by using a special kind of amplifier called an electron multiplier. With such counting, we can determine experimentally the number of electrons passing per second in the electron beam [Fig. 24 – 24(a)].



24–24 **(a) We can measure the number of electrons per second passing by in an electron beam by directing the beam at a counter. (b) We can determine the rate at which the same beam carries charge by firing the beam at a "bottle" with a narrow opening for a long enough time to accumulate a measurable charge on the bottle.**

After we have measured the number of electrons per second in the beam, we can catch the beam in a metal bottle [Fig. 24 – 24(b)]. The electrons which enter the bottle stop in the metal walls, and even if they eject other electrons from the inner surfaces of the wall, hardly any particles will find their way out of the small neck of the bottle. The bottle therefore is charged up only by the charge brought in by the electrons. After a sufficiently long time, the bottle will acquire enough charge to be measured, and we can therefore determine the charge brought into the bottle in each second. By dividing the charge per second by the number of electrons per second in the beam, we then find the charge of an electron. It is exactly one negative elementary charge.

We can also use the same kind of measurement to measure the charge of the particles shot out in the radioactive disintegration of polonium. Indeed, the first experiment of this kind was done by Rutherford and Geiger in 1908 in order to find the charge of these atomic fragments, which we usually call alpha particles. They found that an alpha particle carries exactly two positive elementary charges. Later experiments showed that an alpha particle can be decomposed into more basic units, into two protons and two neutrons. As we already know, the neutrons are neutral. Each of the protons, on the other hand, has one positive elementary charge. A number of other charged particles are now known which differ in mass from electrons or protons. But when we carry the decomposition of matter far enough, all the electric particles found always have either one positive or one negative elementary charge.



24–25 For Question 2.

QUESTIONS

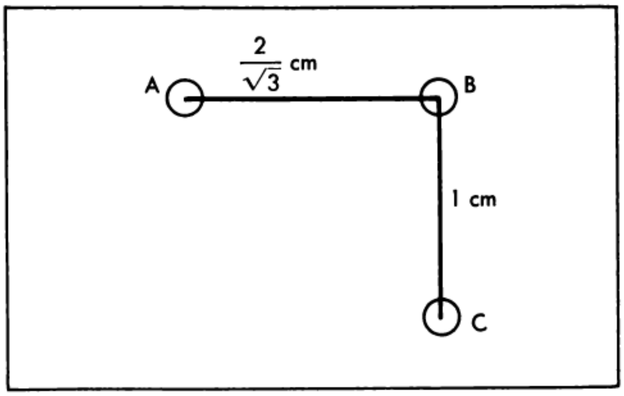
1.* Two small electrified objects *A* and *B* are separated by 0.03 meter, and repel each other with a force of 4.0×10^{-5} newton. If we move body *A* an additional 0.03 meter away, what is the electric force now?

2. Three equally charged objects are located as shown in Fig. 24 – 25. The electric force exerted by *A* on *B* is 3.0×10^{-6} newton.
   (a) What electric force does *C* exert upon *B*?
   (b) What is the net electric force on *B*?

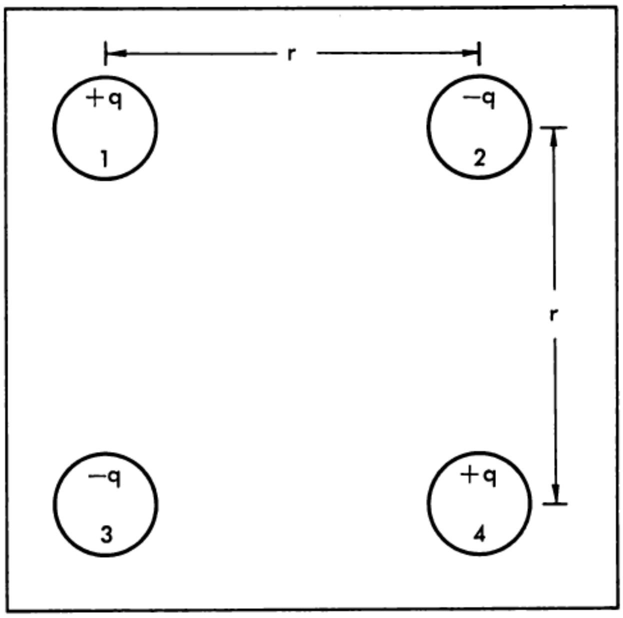
3. Suppose that we place three small charged spheres, with equal charges on them, as shown in Fig. 24 – 26. Sphere *C* exerts a force of 4×10^{-6} newton on *B*.
   (a) What force does *A* exert on *B*?
   (b) What is the net force on *B*?

4. Four identical conducting spheres mounted on Dry Ice pucks are held on a smooth level table at the four corners of a square. Two of them carry a charge of $+q$ and the other two carry a charge of $-q$ each, as shown in Fig. 24 – 27.
   (a) Make a vector diagram, to scale, showing the electric forces on sphere 1.
   (b) If all the pucks are released at the same time, in what direction will each of them move?

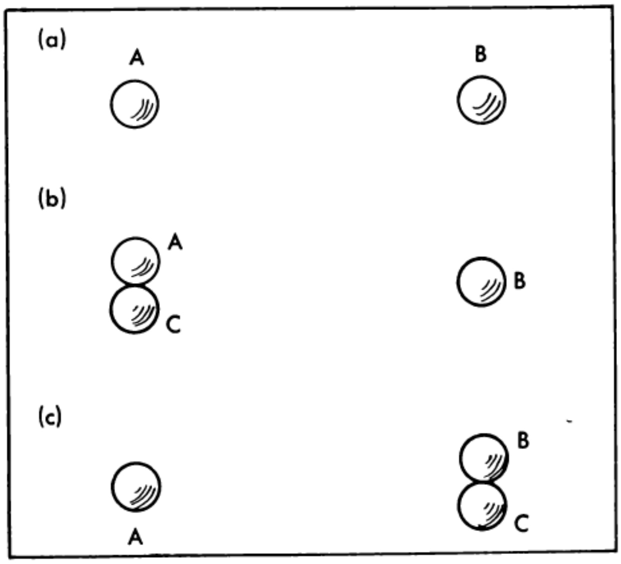
5. Two equally charged identical conducting spheres *A* and *B* repel each other with a force of 2.0×10^{-5} newton [Fig. 24 – 28 (a)]. Another identical uncharged sphere *C* is touched to *A* [Fig. 24 – 28 (b)] and then moved over right next to *B* [Fig. 24 – 28 (c)].
   (a) What is the electric force on *A* now?
   (b) What is the net electric force on *C* (after it has touched *A*) when it is halfway between *A* and *B*?



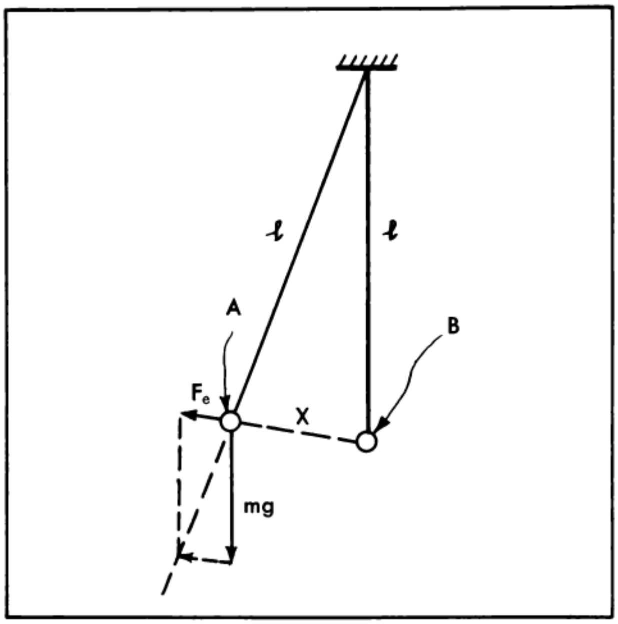
24–26 For Question 3.



24–27 For Question 4.



24–28 For Question 5.



24–29 For Question 6.

6. Two charged balls *A* and *B* each with mass *m* are placed as shown in Fig. 24 – 29. The ball *A* is free to move; *B* is fixed in position.

   (a) How does the force F_e between them depend upon *X*?

   (b) Find the relationship between the electric force F_e and the weight *mg* in terms of *X* and *l*.

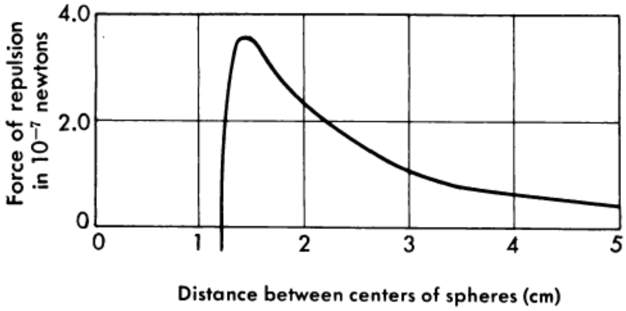
   (c) If the charge remains the same and the separation *X* is cut in half, how must the weight have changed?

7. The graph in Fig. 24 – 30 shows the electric force of repulsion on a tiny charged conducting sphere as a function of its separation from a large conducting sphere. The large sphere has a radius of 1 cm, and has 10 times the charge of the small sphere.

   (a) How is the force changing as the separation changes from 5 cm to 3 cm?

   (b) Explain the behavior of the force between separations 2 cm and 1 cm.

8.* What would happen to the force on the movable charge in Fig. 24 – 5 if charge 3 changed to the opposite sign?



24–30 For Question 7.

9.* Match the fields shown in Fig. 24 – 8 with the grass-seed patterns of Fig. 24 – 9.

10.* Can electric field lines cross one another?

11.* If a charged particle is free to move, will it travel along an electric field line?

12. If you continue the experiment shown in Fig. 24 – 18, what sphere velocities will you expect to get in the situations in Fig. 24 – 31?

13. Suppose we measure the speed of the plastic spheres between the plates of a micro-micro-balance and find it is three halves of their speed under gravity alone. The weight of a sphere is 2.8×10^{-14} newton.

    (a) What is the electrical force on them if they are moving upward?

    (b) If they are moving downward?

14. There is a relationship between the charge on two parallel plates and the force they exert on a charged object between them. There is also a relationship between the number of batteries connected and the force on a charged object between the plates. What is the relationship between the number of batteries and the amount of charge the batteries put on the plates?

15. How would increasing the size of the plastic spheres used in a Millikan experiment affect their terminal velocity?

16. (a) If we have two large parallel horizontal plates with equal and opposite charges, how much does each plate contribute to the total field?

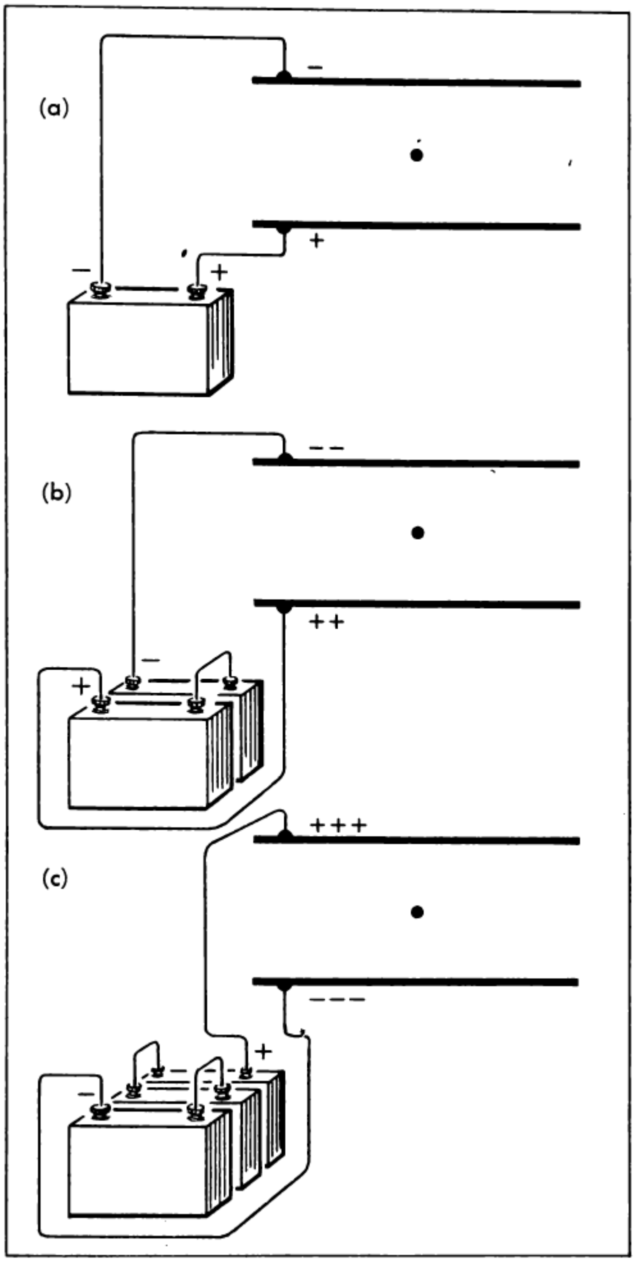
    (b) If the force on a single elementary charge between the plates is 1.4×10^{-14} newton, what is the force on an elementary charge placed above the top plate and close to its center?

    (c) If the force on a single elementary charge between the plates is 1.4×10^{-14} newton, what force does the top plate exert on each elementary charge on the bottom plate?

17.* If the force on each elementary charge in a Millikan experiment was 1.4×10^{-14} newton, what would it be if we increased the distance between the plates by a factor of 10?

18. Refer to Fig. 24 – 22(b).

    (a) Without moving the plates, how can we in-



24–31 For Question 12.

crease the force on the elementary charge between the plates which are a distance $2d$ apart so that it will equal the force on the plates which are a distance d apart?

(b) The plates which are a distance d apart are insulated so that when they are disconnected from their battery the number of charges on the plates remains constant. When the plates are disconnected and then separated to a distance $3d$, it is found that the force on the elementary charge has not changed. If they are reconnected to batteries, how many batteries would have to be connected in series with these plates so that force will not change?

19. By using an argument similar to the one followed in the box on page 446, show that the field at the center of a uniformly charged hemispherical shell is independent of the radius, if the charge per unit area remains constant.

20. Suppose we have a hollow sphere of nonconducting material, which we charge uniformly —say, positively. Assume that the sphere is cut into two parts *A* and *B* by passing a plane through the sphere, as shown in Fig. 24 – 32.

(a) When part *A* is removed, what is the direction of the electric field at the center of the sphere due to part *B*?

(b) If we had removed part *B* instead of part *A*, what would the direction of the field at the position of the center of the sphere have been? How would its magnitude compare with the magnitude of the field produced by *B*?

(c) If the positive charge on *A* is changed to a negative charge of equal amount, how will the field at any point *P* inside the sphere compare with that due to *B* alone?

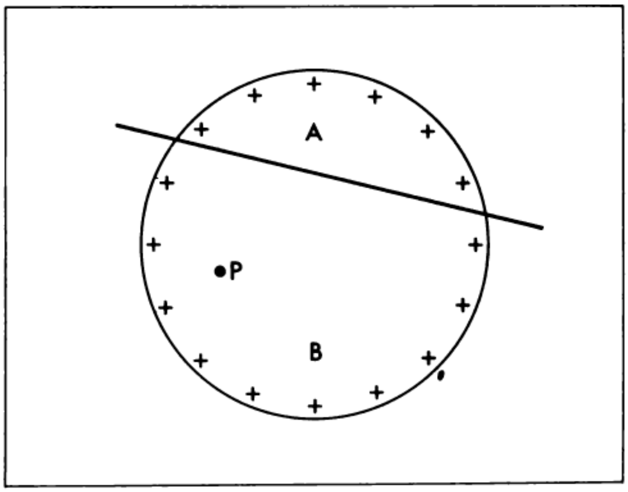
21. Only one of several identical metal spheres is charged. The charged sphere *A* experiences a force $F = 1.0 \times 10^{-4}$ newton when placed midway between certain charged parallel plates.

(a) Sphere *A* is then touched to one of the uncharged spheres (*B*). What force does *A* now experience when placed between the same plates?

(b) *A* is touched to another of the uncharged spheres (*C*). What force does *A* now experience?

(c) What force is exerted on sphere *B* when it is between the plates? On sphere C?

22. Two large metal plates are separated by 0.10 m. They are connected to the terminals of a battery. A small charged ball halfway between them experiences an electric force of 3×10^{-4} newton.

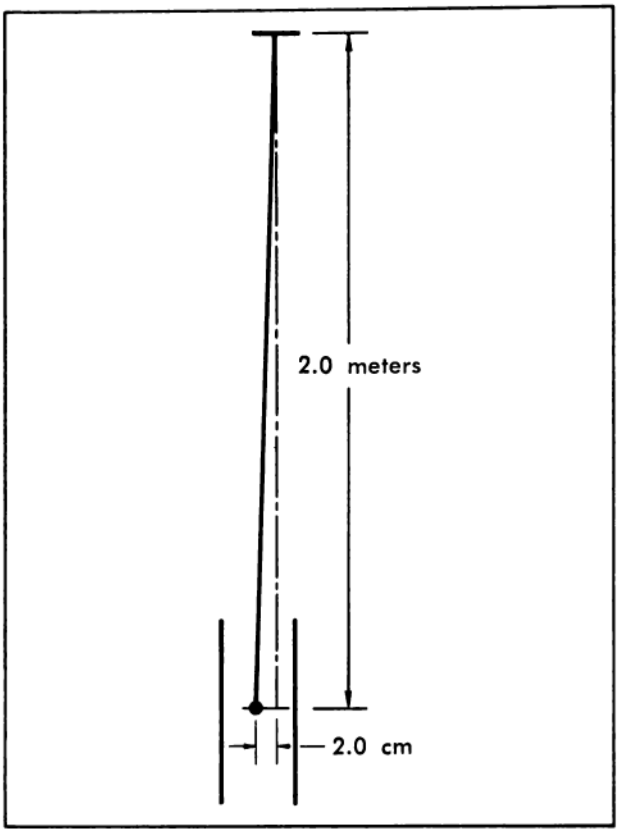


24–32 For Question 20.

The plates are now moved apart until the separation is 0.15 m. The same battery is connected to them.

(a) What force now acts on the ball?

(b) If we add two more identical batteries in series at the new separation, what is the force on the ball?



24–33 For Question 25.

23.* What will be the force of repulsion between two small spheres 1.0 meter apart if each has an excess of 10^6 elementary charges?

24. A plastic sphere of mass 3.06×10^{-15} kg is in an electric force field which exerts an upward force of 1.00×10^{-14} newton on each positive electric particle. The electric force exerted on the sphere is sufficient to balance the force of gravity on it.

(a) What is the excess of electric particles on the sphere?

Suppose that each molecule has a mass of 3.00×10^{-26} kg.

(b) How many molecules are there in the sphere?

(c) What fraction of the molecules in the sphere have lost or gained an elementary charge?

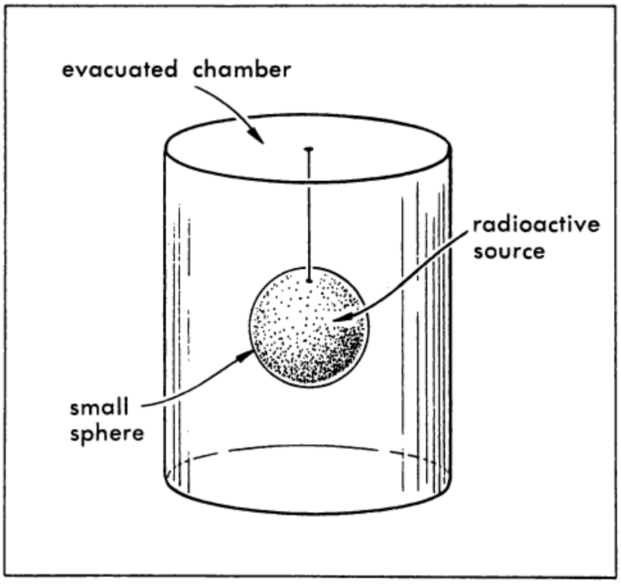
25. A sphere of mass 4.5×10^{-3} kg is hanging on a string 2.0 meters long between two oppositely charged parallel plates as shown in Fig. 24–33. At equilibrium the ball has been pulled 2.0 cm from its original position.

(a) What is the magnitude of the electric force on the sphere?

The electric force on each elementary charge between the plates is 3.0×10^{-14} newton.

(b) What is the excess of electric particles on the ball?

26. In an experiment two small conducting spheres (120 milligrams each) were suspended on two fine insulating fibers 82 cm long. When equal charges were placed on the two spheres, they separated and came to rest with a distance of 10 cm between centers. What was the number of elementary charges on each sphere?



27–34 For Question 27.

27. A thin layer of radioactive material emitting beta particles is painted on a small conducting sphere. The sphere is then hung on an insulating thread in an evacuated chamber (Fig. 24–34).

After 50 days we measure the charge on the sphere and find it to be 4.32×10^{10} positive elementary charges. (The sphere has been kept in an evacuated chamber, so very few charges have leaked off.) In a separate experiment with a Geiger counter we find beta particles shooting out at the average rate of 1.00×10^4 per second.

(a) What is the charge of one beta particle?

(b) What do you think beta particles may be?

25

Energy and Motion of Charges in Electric Fields

THIS CHAPTER deals with the motions of charges in electric fields. We shall use these motions to determine the masses of charged particles and to discuss the currents in electric circuits. In particular we shall find the masses of positive hydrogen ions and of electrons, because these charged particles will be central to our later study of atoms. When charged particles accelerate, their kinetic energy changes. Energy is transferred to the particles. The energy transfers (from chemical stored energy, to the energy of moving charges, to thermal energy) will allow us in the next chapter to consider all kinds of electric circuits from a single point of view.

25–1 DETERMINING THE MASS OF THE ELECTRON AND OF THE PROTON

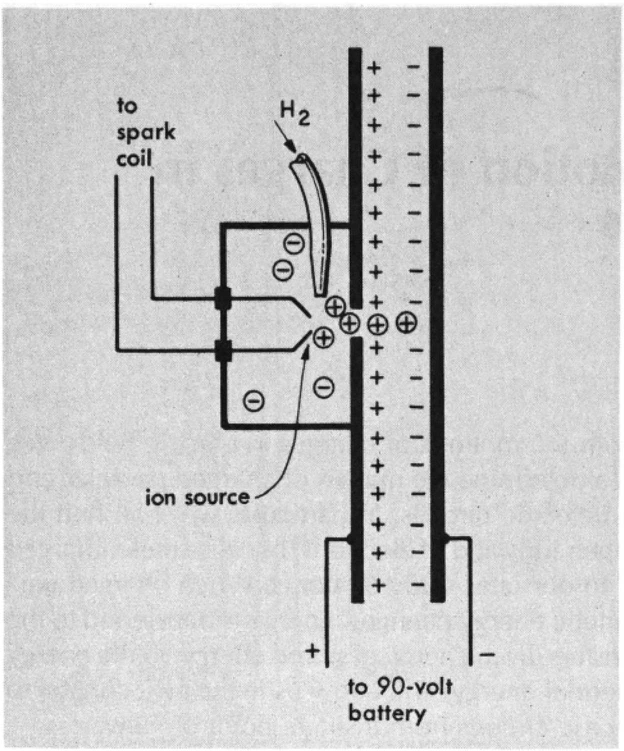
In this section we shall show how the mass of an electron or proton can be determined experimentally by accelerating the charged particle through a known uniform force field a known distance and measuring its speed. As we have learned, if an object moves a distance d in the direction of a force $\vec{\mathbf{F}}$, the work Fd done on the object is equal to the gain in kinetic energy of the object. Now, if the object starts from rest, this is also the final kinetic energy of the object:

$$Fd = \tfrac{1}{2}mv^2$$

Hence, if we know F, d, and v, we can determine the mass m.

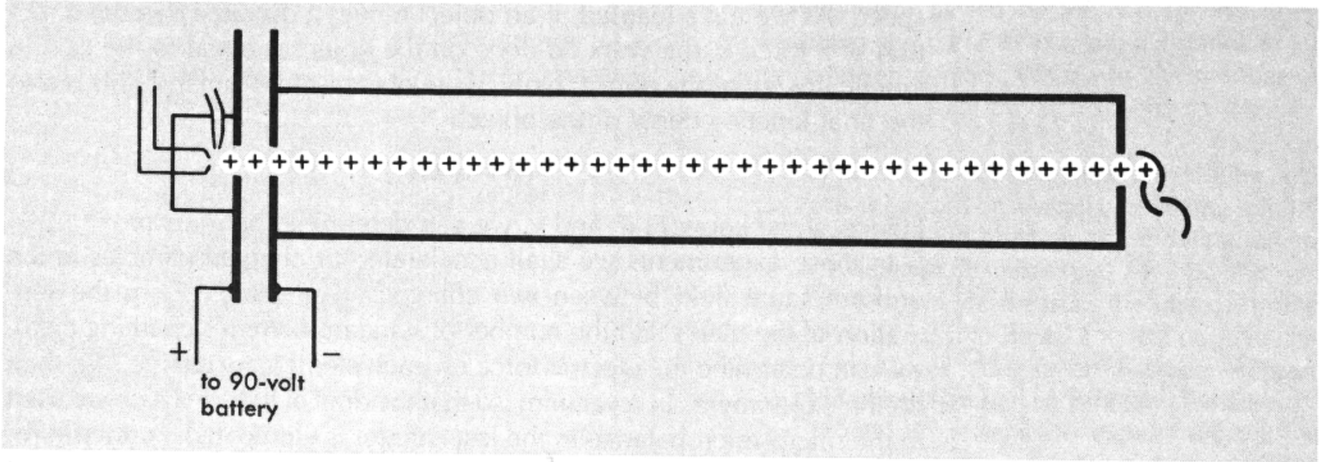
In these experiments we shall accelerate our charged particles in the uniform force field between two charged metal plates. From the separation of the plates and the number of standard batteries charging them, we can determine the electric force on each elementary charge. We shall do the experiments in a vacuum, so that the drag of the air which we used in the micro-microbalance in the last chapter is eliminated. Furthermore, because the protons and the electrons are so light, with masses less than 10^{-11} times the mass of the standard spheres we used in the micro-microbalance, we can neglect the gravitational forces in comparison with the electric forces in these experiments.

We ionize some hydrogen near a pair of charged plates and let some of the ions drift with negligible speed through a small hole into the region between the plates (Fig. 25–1). The electric force field between the plates accelerates the ions as they cross from one plate to the other, giving them a final kinetic energy $\frac{1}{2}mv^2$. Now, in the right-hand plate there is a small hole through which some of the ions can pass into a chamber 0.50 meter long (Fig. 25–2). This chamber is made of a conductor, and since there is no electric force inside it, the ions move its entire length without changing their velocity. It takes the ions only a few microseconds (a few times 10^{-6} second) to go the whole distance. Although this time interval is very short, with a special timer we can measure it accurately. Then we can find an accurate value for the speed v of an ion.

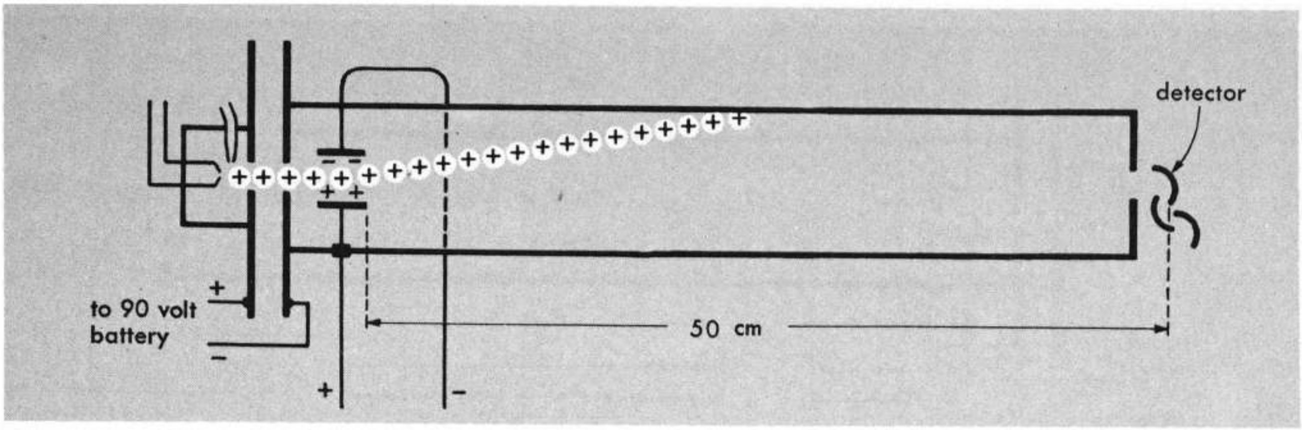


25–1 Apparatus for accelerating hydrogen ions in an electric force field. The ions are formed in the box at the left, and some drift through the hole into the region between the plates. The whole setup is enclosed in an evacuated chamber.

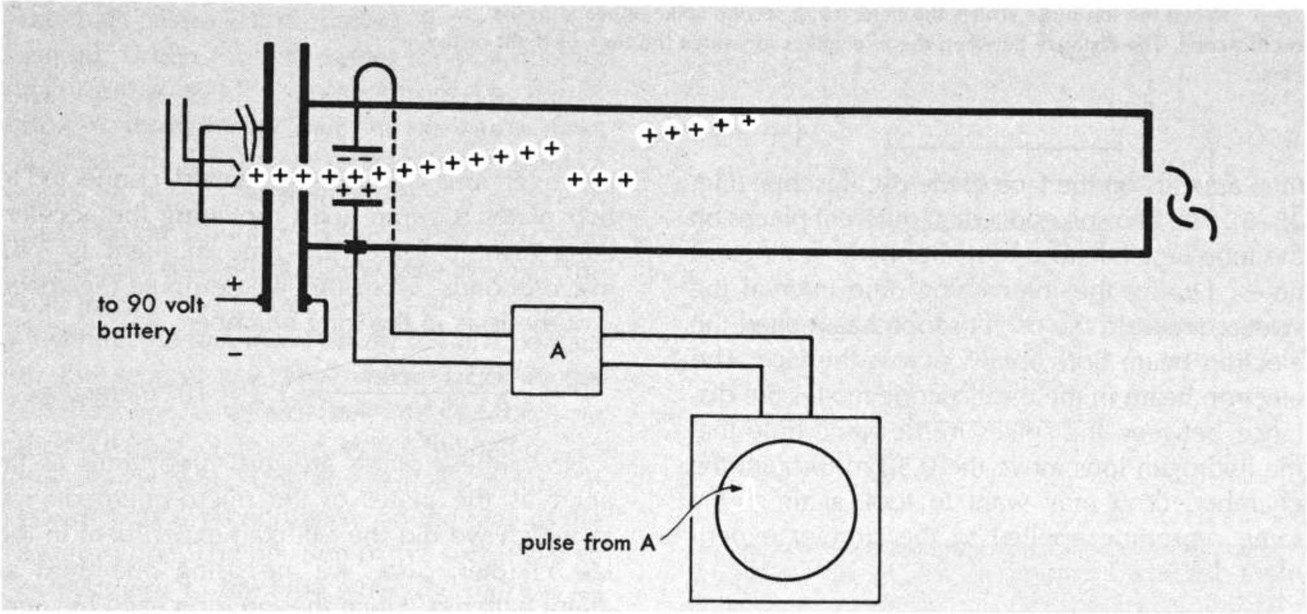
In order to measure the time it takes for the ions to pass from one end of the long chamber to the other, we must record the time at which a particular ion passes a given point at the left and the time at which the same ion reaches the far end on the right. (You can understand the idea of such a measurement by considering the experiment discussed in the box, page 465). To mark the time when a given ion enters the long chamber, we place a couple of small deflecting plates near the entrance (Fig. 25–3). With these deflecting plates we can control the direction of the beam of hydrogen ions. When the deflecting plates are charged, the hydrogen ions are subject to a sidewise electric force which bends them out of their normal path. If the deflecting plates are then discharged, only the ions just



25–2 To measure the mass of hydrogen ions, we drill a hole in the right-hand plate, allowing some ions to enter a long chamber on the right, where they move without forces.



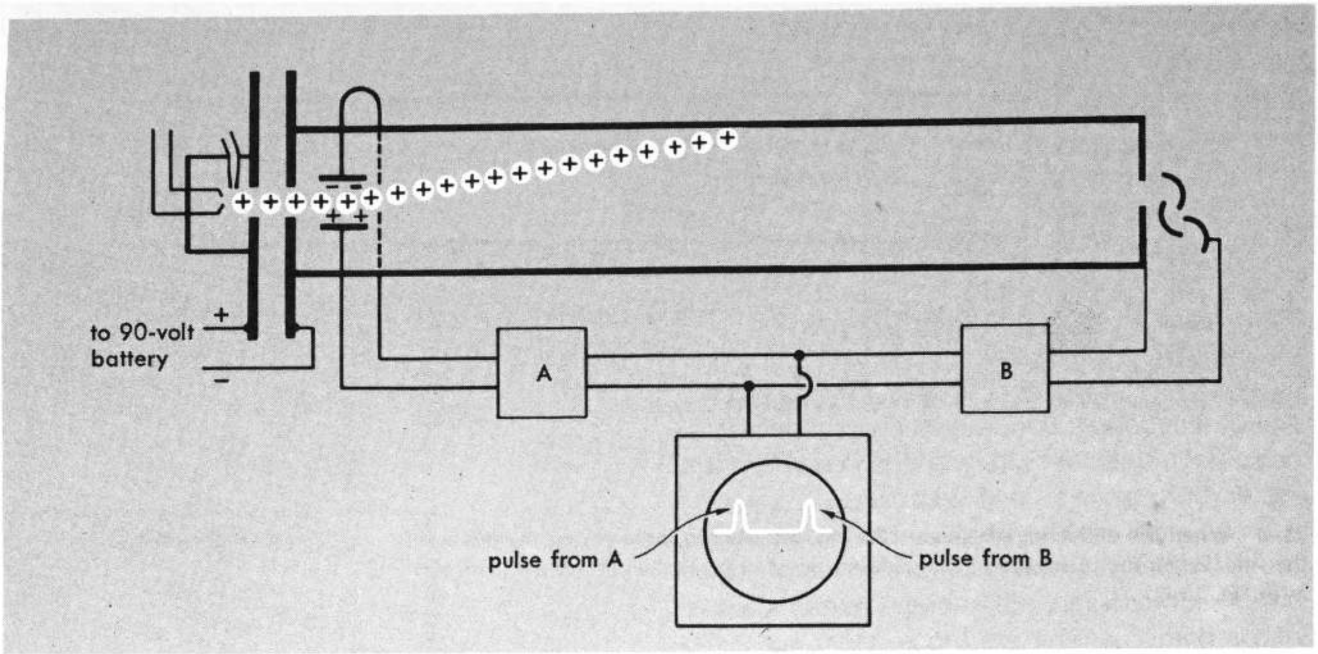
25–3 **When the deflecting plates are charged, the electric force swings the beam to the side. When the plates are uncharged, ions go straight ahead and hit the detector, as in Fig. 25–2.**



25–4 **When we momentarily discharge the plates, we let through a short pulse of ions that goes straight along the chamber to the detector. The discharge is recorded as a spike on the oscilloscope, marking the time that the pulse begins its flight.**

entering the chamber have a velocity along the long axis; hence, the first ions to come out through the hole at the far end will be those which have traveled the entire 0.50 meter during the time since the plates were discharged. The arrival of these ions is signaled by a detector placed just outside the hole.

To measure the time interval between the time of discharge and the arrival of the first ions at the detector, we connect the deflecting plates in the chamber to the vertical deflecting plates of an oscilloscope (Fig. 25–4). When we discharge the plates in the long chamber, the time of the discharge is recorded as a spike in the curve traced on the face of the oscilloscope. We also connect the detector at the far end of the long chamber to the same vertical deflecting plates of the oscilloscope. (We make the electrical connections at both ends of the chamber identical.) When the beam of ions hits the detector, a second spike



25–5 When the ion pulse strikes the detector, a second spike appears on the oscilloscope. The distance between the two spikes measures the time of flight of the ions.

then appears on the face of the oscilloscope (Fig. 25–5). The two spikes occur at different places on the tube face because they are made at different times. During the intervening time interval the sweep circuit in the oscilloscope has moved the electron beam horizontally across the face. The electron beam in the oscilloscope moves the distance between the spikes in the same time that the hydrogen ions move the 0.50 m through the chamber. (You may want to look again at the same reasoning applied to the simpler experiment described in the box.)

With a modern oscilloscope, the sweep circuit can move the electron beam horizontally across the face of the tube from one side to the other in a few hundredths of a microsecond. To measure the speed of the ions, we adjust the sweep circuit to make the whole trace on the oscilloscope represent 5 microseconds. The two spikes on the trace are then well separated on the face of the oscilloscope. By measuring the distance between the spikes, we measure the time it takes the beam to cross the long chamber. We obtain the time interval from the moment when the beam is allowed to go straight ahead to the moment when it hits the detector, with an accuracy of about one-hundredth of a microsecond. With hydrogen ions and a 90-volt battery connected to two plates 9.3 mm apart providing the accelerating electric force, the time of flight is 3.82 microseconds. From this we compute the speed v of the ions in the long chamber:

$$v = \frac{0.50 \text{ meter}}{3.82 \times 10^{-6} \text{ sec}} = 1.3 \times 10^{5} \text{ meters/sec}$$

Now, these plates are just three times as far apart as the plates of the micro-microbalance in which we did the Millikan experiment in the last chapter; also, we are using one third as many batteries. Since the force on each elementary charge is proportional to the number of batteries and is inversely proportional to the separation of the plates, we must have one ninth the force on each elementary charge, or $\frac{1}{9}$ (1.4×10^{-14} newton) on each elementary charge. If we assume that the hydrogen ion carries one elementary charge, each ion experiences a force this great when it is between the plates. As the ion goes from one plate to the other, it travels a distance 9.3×10^{-3} m in the direction of the force, so the work done on it is

$$Fd = \tfrac{1}{9}(1.4 \times 10^{-14} \text{ newton})\,(9.3 \times 10^{-3} \text{ m})$$
$$= 1.4 \times 10^{-17} \text{ joule}$$

Hence,

$$\tfrac{1}{2}mv^2 = \tfrac{1}{2}m(1.3 \times 10^5 \text{ m/sec})^2 = 1.4 \times 10^{-17} \text{ joule}$$

and solving for m, the mass of the hydrogen ion, we find

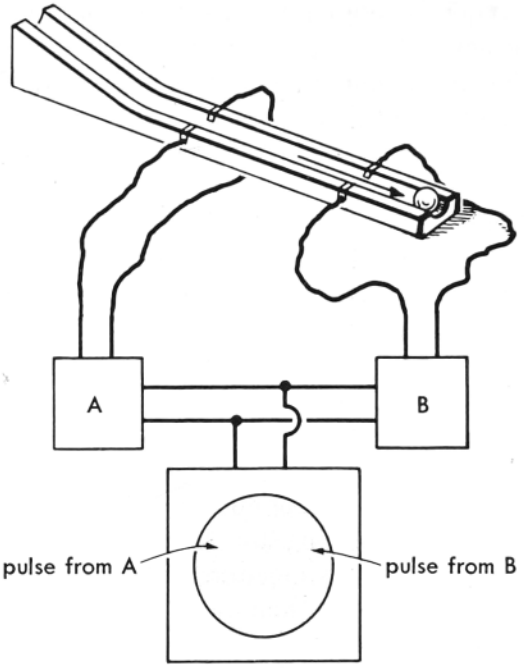
$$m = 1.7 \times 10^{-27} \text{ kg}$$

We can now state our result. If the hydrogen ion is singly charged, we find it has almost the same mass as the hydrogen atom. If fact, we can go one step further and assert that the hydrogen ion actually does carry a single charge and does have practically the mass of the atom. This must be true, because the assumption that the ion carries more charge leads to an absurd result. For example, if an ion carried two elementary charges, the actual value of $mv^2/2$ would be just

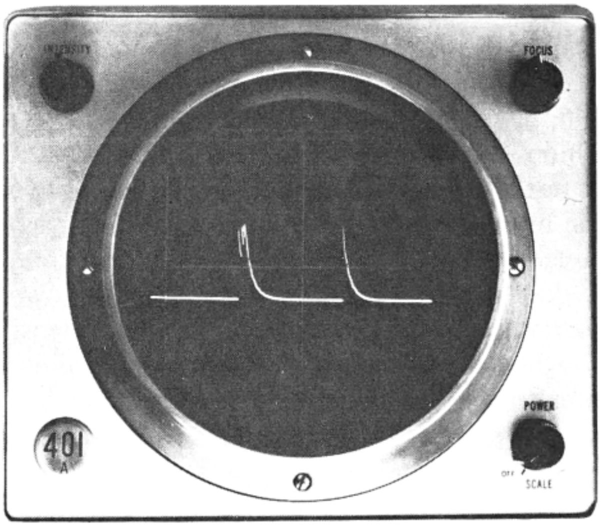
A Measurement of Speed by Time of Flight

We can understand the idea of this experiment by measuring the speed of a metal ball moving steadily along a trough. In Fig. 25–6 the trough is shown. At two points on the ball's path there are contacts in the trough; and as the ball passes each contact, it completes a circuit. When the ball passes the first contact a current flows for a moment from a battery in Box *A* through the ball. At the same time, triggered by the flow of charge, Box *A* sends a pulse to the oscilloscope so that a spike appears on the oscilloscope face. The horizontal position of this spike tells the time at which the ball passes the contact.

While the ball moves from the first contact to the second one, the sweep circuit of the oscilloscope continues to trace the curve on the face from left to right. When the ball passes the second contact, an identical mechanism in Box *B* puts a second spike on the face of the oscilloscope. The position of this second spike corresponds to the time when the ball passes the second contact. The distance between the two spikes measures the time interval, the "time of flight" of the ball, between the first and second contacts in the trough. Figure 25–7 shows a record plotted out on the face of an oscilloscope in such a measurement. The sweep traced the curve steadily, moving from one side of the grid of lines on the face of the oscilloscope to the other in 1 second. So the distance between spikes corresponds to 0.38 sec. In this time the ball moved 24 cm from contact to contact. So its speed was 0.63 meter per second.



25–6



25–7

twice what we used above. Since we measured v, this could only mean that the mass of the ion is twice what we found. Such a hydrogen ion would have twice the mass of the atom of which it is a fragment. And that conclusion is unlikely enough that we reject it.

We have already had indications that electrons are small building blocks common to all atoms. Presumably a hydrogen ion is a hydrogen atom that has lost an electron. Furthermore, we never see a positive fragment of hydrogen with two positive elementary charges in either these or any other experiments. This is one of many indications that the positive hydrogen ion is an elementary building block. It is the proton. And, as we mentioned in Chapter 24, protons are also found in alpha particles and fragments of heavier atoms.

When hydrogen is split into charged fragments, as we just found out, the proton has almost the whole mass of the atom. The electron therefore must be very light. We can use the same apparatus to measure the electron mass and thus verify this conclusion.

If hydrogen ions are made by taking an electron out of a hydrogen atom, the source of ions at the left of our apparatus should also provide electrons. By turning the battery around, we can accelerate the electrons that drift through the hole in the left-hand plate and reject the hydrogen ions. (Alternatively we can use the electrons boiled out of a hot filament—with either source we get the same results.) Since the electrons are pushed across between the plates with the same force through the same distance, they will gain the same value of $mv^2/2$, that is,

$$mv^2/2 = 1.4 \times 10^{-17} \text{ joule}$$

Because they are much lighter than the positive ions, they will be moving faster when they pass through the hole in the second plate and cross the long chamber. Consequently, when we measure the speed with which the electrons pass through the long chamber, we find that we are closer to the limits imposed by our apparatus. With the sweep circuit adjusted so that the distance across the tube face corresponds to a time interval of about 0.12 microsecond, we can see the two spikes recorded on the face of the oscilloscope. They are about nine one-hundredths of a microsecond apart. However, to measure the exact time interval between them to better than 10 percent accuracy requires an unusually fast oscilloscope and very sharp spikes. Taking 0.090 microsecond as a first approximation, we find that the speed is about 5.6×10^6 meters per second. Consequently, the mass of an electron (computed in the same way as the mass of positive ions) is about 0.9×10^{-30} kg. Accurate measurements of the mass of the electron give 0.911×10^{-30} kg, about a two-thousandth of the mass of a hydrogen atom.*

In measuring the masses of the ions, we can change the details of the experiments so that we use different known electric forces and different separations between the plates. We then get different values of $mv^2/2$. But we always get the same masses. We can use other methods to measure the masses—and we shall describe another one in Chapter 28. The results are the same. Indeed, the masses of many ions are now extremely well established, and they lead to accurate masses of the atoms.

The experiments we have just described tell us about two of the fundamental building blocks from which atoms are made, protons and electrons. In interpreting these experiments, we naturally used the Newtonian mechanics that we studied earlier. Because our results are sensible and consistent, we may now conclude that Newtonian mechanics can be applied over an enormous range. It gives an accurate description of the motions of large masses at large distances, of the planets in the solar system; it applies to moderate masses at the moderate dimensions of everyday life; and it also applies to the small masses of electrons moving through relatively small distances at huge speeds. Consequently, it is reasonable to try to apply the same mechanics and the same forces to the electrons within atoms. In particular, we can make a model of a hydrogen atom in which a single electron moves around the relatively massive proton under the influence of the Coulomb force. Because the Coulomb force has the same inverse-square dependence as the gravitational force between the sun and a planet, we shall call this model a planetary model of the atom. In the planetary model of hydrogen, the electron moves around the proton in an elliptical orbit like a planet around the sun. The planetary model, as we shall see in the next chapter, meets with some difficulties, but it is a good first approximation to actual atomic behavior. Since Rutherford introduced it in 1911, it has played a central role in our understanding of atoms.

---

*With modern fast oscilloscopes and amplifiers, we can measure time intervals to within 10^{-9} sec. Therefore it is now possible to use this technique to get an accuracy of about 1 percent. Earlier, precision measurements were made by other methods which we shall come to later on.

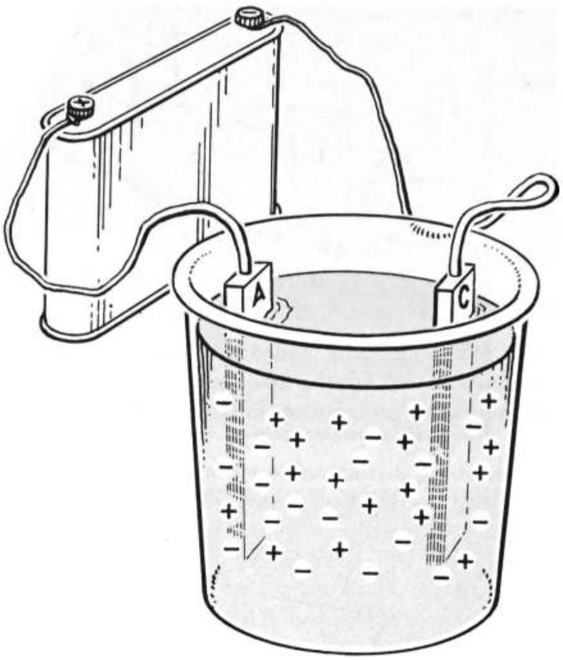
25-2 ELECTRIC CURRENT

The beams of protons and electrons in the experiments described in the last section are electric currents. The natural measure of these currents is the number of elementary charges transported per second.

If we catch a beam of positive particles in a metal cup, the cup becomes more and more positively charged as time goes on. For this reason we say that the direction of this electric current is into the cup. The current is in the same direction as the motion of the positive particles.

A beam of negative particles is also an electric current. But when we catch it in a metal cup, the charge of the cup decreases. If we catch beams of positive and of negative particles at the same time, they neutralize each other. When the same number of positive and of negative particles arrive in the cup per second, the charge of the cup does not change. There is no incoming electric current. Even though the beam of negative particles moves *toward* the cup, it constitutes an electric current *away* from the cup, emptying the cup of positive charge as fast as the positive beam fills it. In other words, a beam of negative particles constitutes an electric current in the direction opposite to the motion of the negative particles.

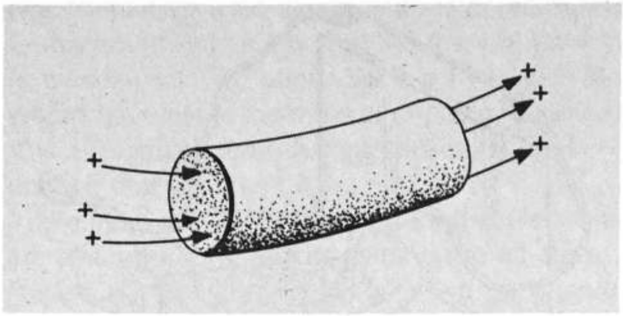
In general an electric current can be brought about by the motion of both positive and negative particles. The total electric current is made of the number of positive elementary charges passing per second (contributing to the electric current in their direction of motion) *plus* the number of negative elementary charges passing per second (contributing to the electric current in the direction opposite to their motion).



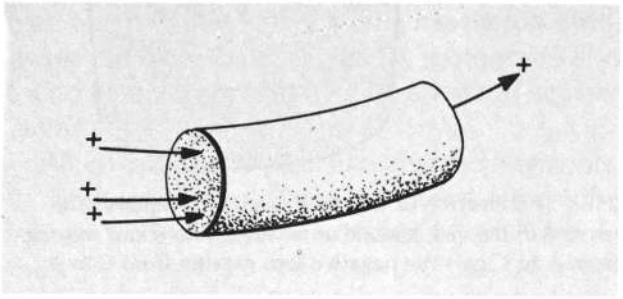
25-8 An electrolytic cell in series with a battery. The current in the cell is made up of the positive ions moving from A to C plus the negative ions moving from C to A.

You already know an example of an electric current made of charges of both signs moving in opposite directions. In an electrolytic cell such as that shown in Fig. 25 – 8, ions move in both directions. The positive ions moving from *A* to *C* are part of an electric current toward *C*. The rest of that electric current is given by negative ions moving in the opposite direction.

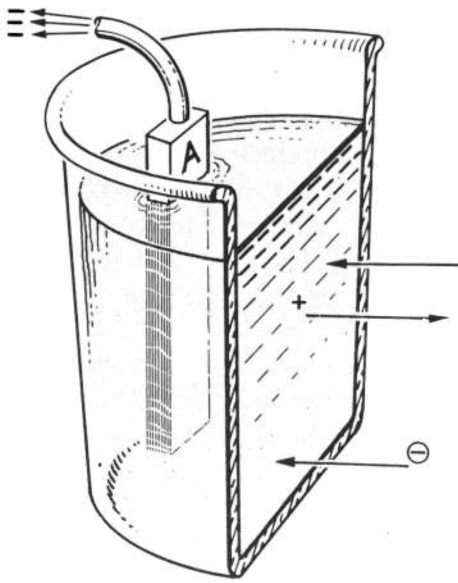
The total electric current does not begin or end in the electrolytic cell. It comes in through the wire from the positive battery terminal to *A* and it goes out through the wire from *C* to the negative battery terminal. In each of these wires the moving charged particles are electrons, and these negative particles move in the opposite direction from the electric current. For example, they go from *A* to the positive terminal, causing the effective transport of positive charges from the terminal to *A*. Looking at the whole circuit, we see that the electric current is flowing from the positive to the negative battery terminal around the external circuit. Inside the battery, on the other hand, the current goes from the negative to the positive battery terminal. The battery moves more positive charges to the positively charged terminal and negative charges to the negative terminal despite the electric repulsions. In this process the



25–9(a) **When the same number of elementary charges goes in one end and out the other end of a section of a circuit, no charge accumulates in that section.**



25–9(b) **If more elementary charges go in than go out, charge accumulates in the region. Therefore, the observation that charge does not accumulate shows that the current into one end of the region must equal the current out the other end.**



25–9(c) **In the region of the circuit shown, electrons flow through the wire and ions move in the electrolytic cell. Because no charge accumulates, we know that the sum of negative and positive ions passing through the center of the cell must equal the number of electrons passing through the cross section of the wire.**

battery uses chemical energy to force the charges onto the terminals.

The net rate of flow of elementary charges is the same all around the circuit. This must be true because, when the current is running steadily, charge does not accumulate in any region of the circuit. The same number of elementary charges, therefore, goes in one end of any region as comes out the other (Fig. 25 – 9).

The steady electric current cannot be measured by observing changes in charge; but we may imagine an observer who can count elementary charges passing any particular point in the circuit. If he watched the charges passing through the wire at a particular place, he would find *I* negative elementary charges, *I* electrons, passing per second. If he looked at the charges passing through the electrolytic cell at a particular place, he would also see *I* elementary charges passing. But here he must count the negative ones moving one way and the positive ones moving the opposite way and add the two numbers in order to get the total electric current which measures the net transport of charge (Figs. 25 – 9 and 25 – 10). In the next section we shall see how a real measurement of a steady current can be carried out.

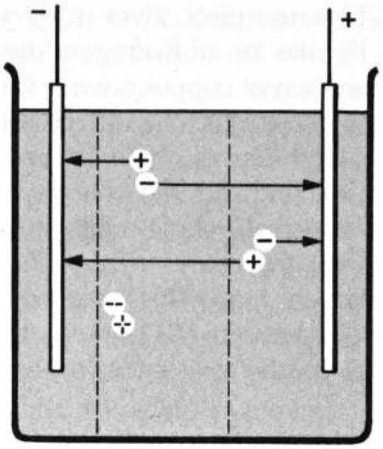
25–3 ELECTROLYTIC MEASUREMENT OF ELECTRIC CURRENTS

The observer who can see elementary charges passing is imaginary. But in certain circumstances we can make equivalent observations. For instance, we have already counted electrons in a weak beam from an electron gun. In this section we shall describe the measurement of the number of elementary charges by observing the amounts of mass transported by ions in solution.

For this purpose we connect a number of electrolytic cells *in series* with a battery as shown in Fig. 25 – 11. With this arrangement, the same amount of charge flows through each cell. The various cells contain solutions of salts of different metals, for example silver and copper. The electrodes are made of platinum (or carbon) which does not react chemically with either of these solutions. We also include a cell filled with dilute acid such as hydrochloric acid (HCl), whose molecules contain hydrogen instead of a metal, and a complicated cell containing aluminum in the compound Al_2O_3. The electrolytic cells for copper, silver, and hydrogen ions are ones that you can handle. The cell for the aluminum ion is different. Aluminum oxide, Al_2O_3, is dissolved in molten cryolite, and the aluminum which goes to the negative electrode is also molten. This kind of cell is actually used in refining aluminum.

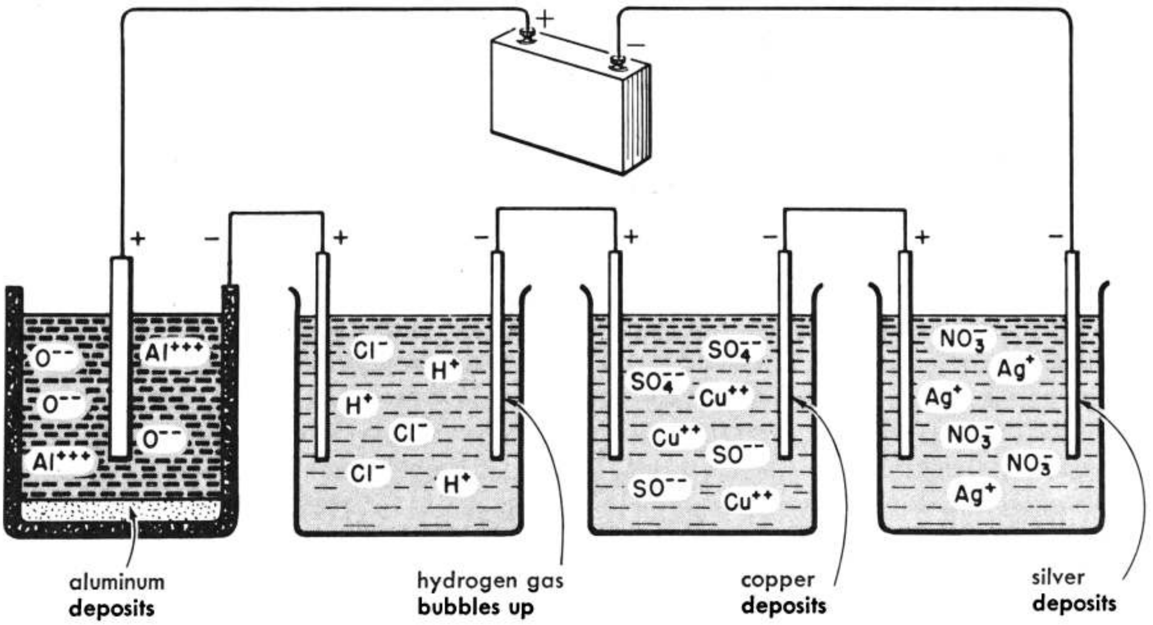
As ions move through the solutions, the negative electrodes become coated with the metals present in the dissolved salts. Also, hydrogen gas bubbles out at the negative electrode of the cell full of acid. This shows that when the molecules of the salts are ionized, the positive ions carry the metal atoms with them (or perhaps they *are* positively charged metal atoms). When these ions arrive at the negative electrodes, the metal atoms usually remain attached to the electrodes. It also shows that the hydrogen ions are positive, and when they are neutralized at the negative electrode hydrogen gas is formed.

By weighing the electrodes before and after an electric current has passed through the cells, we can determine the mass of metal that has been deposited in each cell. From this mass and from the mass of a mole of the metal we can compute the number of moles deposited. Similarly, we can determine the number of moles of hydrogen. When these computations are made, a remarkable fact appears. For each mole of silver deposited there is a mole of hydrogen atoms, half a mole of copper, and one third of a mole of aluminum.



25–10 When any number of ion pairs is produced in the cell, this number of positive ions will be collected at the negative electrode, and this number of negative ions will be collected at the positive electrode. The arrows show how the ions go to the electrodes. Count the arrows passing through the dotted boundaries. The total number of positive and negative ions crossing any boundary between the electrodes is the same. Try adding the pairs produced at any place you want.

Since a mole of any element contains the same number of atoms, we see that the charge



25–11 Electrolytic cells in series with a battery.

carried by one ion of silver is the same as that carried by one ion of hydrogen; the charge carried by one ion of copper is twice that carried by one ion of silver; and the charge carried by one ion of aluminum is three times as large. In general, we find that the charges of all ions in solution are small whole-number multiples of a unit charge, the charge of the silver ion (or of the hydrogen ion). This great discovery was made by Faraday in 1833. It was the first cogent evidence of the existence of an elementary charge.

The smallest charge carried by an electrolytic ion, the charge of the silver ion or hydrogen ion, must be equal to the elementary charge we found in the Millikan experiment in Section 24 – 5. There are a number of reasons for being sure that this is so. For example, when an electrolytic hydrogen ion moves to the negative electrode in an electrolytic solution, it is made neutral by an electron from the platinum. Then the neutral hydrogen atoms combine to form hydrogen gas which bubbles out at the negative electrode. Because the platinum supplies ordinary electrons—there are no other—and because the hydrogen gas released is the same as any other hydrogen, the hydrogen ions in solution must be the same as the hydrogen ions whose mass we measured in Section 25 – 1. Like those hydrogen ions, these ions must carry exactly one positive elementary charge which can be neutralized by the negative elementary charge of the electron.

At the other electrode in the same electrolytic cell, negative ions are neutralized and neutral gas molecules are formed. The negative ions lose their negative charge by giving up ordinary electrons to the positive electrode; these electrons then flow through the wire toward the positive terminal of the battery.

In general the ions of any kind which come to the electrodes in any electrolytic cell are turned into neutral atoms or molecules, and this neutrality is brought about by combining with ordinary electrons from a platinum electrode or by supplying ordinary electrons to the electrode. All the ionic charges must therefore be multiples of the elementary charge, and presumably the smallest ionic charge is the same as the elementary charge we detected in the Millikan experiment.

By itself the kind of evidence we have mentioned is reasonably convincing. There is no reason to assume that there are two kinds of unit charges in the universe; and the unit charge found in electrolysis seems to fit together perfectly with the unit charge found in the Millikan experiment. We shall be able to give another check on this conclusion in the next section. Furthermore, by the end of this part of the book we shall have plenty of evidence to show that atoms are made of electrons and of positively charged particles which exactly neutralize them. Since we already know that each electron carries one negative elementary charge, it then follows that anything built of atoms must either be neutral or carry a whole number of elementary charges. Because everything, including ions, is so built, we again conclude that the ionic charges are multiples of the elementary charge.

We now know that hydrogen ions and silver ions each carry one positive elementary charge; each copper ion carries two positive elementary charges, and each aluminum ion carries three positive elementary charges. (Also, the charges of some negative ions are indicated in Fig. 25 – 11.) We can use this information to measure the total amount of charge transported across an electrolytic cell. For instance, when only hydrogen ions come to the negative electrode, the number of elementary charges transported is the same as the number of hydrogen ions that arrive. Consequently, by measuring the mass of hydrogen released and dividing by the mass of one atom, we find the number of elementary charges.

We can use a similar method to measure a current. For instance, by putting an electrolytic cell containing silver ions in a circuit and measuring the *rate* at which silver is deposited, we can measure the electric current in elementary charges per second. Since one elementary charge is transported for each silver atom deposited, the electrolytic cell becomes an electric current meter. Indeed, electrolytic cells depositing silver are often used as standards to measure currents. Usually, however, it is more convenient to employ a different form of meter. We can use any device which responds differently and reproducibly to different currents, but we must calibrate at least one of these other meters. We calibrate any other meter by putting it in series with the electrolytic cell. Since the same current passes through both meters, we can mark the scale of the other meter with the values of the current in elementary charges per second measured by the rate of deposit of silver in the electrolytic cell.

25–4 EXPERIMENTAL CHECKS ON ENERGY TRANSFER, ELECTRIC FORCES, AND ELEMENTARY CHARGES

In Section 25 – 1 we used a known electric force $\vec{F}$ to accelerate ions across a distance d. The ions started from rest and acquired the final speed v. We used the relation

$$Fd = \frac{1}{2}mv^2$$

to predict their kinetic energy after the acceleration. Now that we can measure an electric current or a total charge transported in terms of elementary charges, we can do an experiment which shows that Fd correctly predicts the kinetic energy $\frac{1}{2}mv^2$ that the ions gain.

For this experiment we use electrons boiled off a hot wire. We accelerate them in a known electric force field between two parallel metal plates, just as we did in measuring the mass of an electron in Section 25 – 1. The electrons which are accelerated from the negatively charged plate crash into the positively charged plate and lose their kinetic energy. This energy is transformed into thermal energy, heating the plate. With reasonable currents the heating of the plate can be great enough so that an appreciable rise in temperature occurs within a few seconds. By measuring this temperature rise we can determine the amount of energy transferred. And by measuring the current we can determine the number of elementary charges. We thus obtain the kinetic energy of each elementary charge. The results of these experiments confirm our prediction that the kinetic energy of each elementary charge is given by the electric force F acting on an elementary charge times the distance d through which the charge accelerates.

It may seem that this verification that Fd gives the kinetic energy is unnecessary. Indeed there is no surprise in our result, but we can draw two valuable conclusions from it. In the first place, when we computed the work Fd which measures the transfer of energy into the kinetic energy of motions of the electrons, we used the value of the force F on each elementary charge that we had already measured on elementary charges at rest or in slow motion in the micro-microbalance experiment mentioned in Section 24 – 7. Up to now we have assumed without experimental evidence that this same force acts on a fast-moving elementary charge. The experiments we have just described show that the electric force on a moving charge is actually independent of its speed. Because electrons accelerated between the charged parallel plates of Section 25 – 1 attain a speed of about 10^7 m/sec, we see that the electric force on an elementary charge is the same at any speed between zero and this enormous value.

Secondly, these experiments show that the unit charge found in electrolytic experiments is the same as the elementary charge found in the Millikan experiment. We measure the currents with meters which are calibrated against electrolytic cells. The number of charges which transfer their kinetic energy into heat at the positive plate is therefore given as the number of electrolytic unit charges. Consequently, our experiments measure the kinetic energy of each electrolytic unit charge. On the other hand, we predict kinetic energy in terms of the force in the Millikan experiment. Consequently, the kinetic energy predicted is the kinetic energy of each Millikan elementary charge. The agreement between the predicted and the experimental results now shows that the electrolytic unit charge and the elementary charge found in the Millikan experiment must be the same. There is only one elementary unit of electric charge.

Measurements like those we have discussed in this section were performed in 1897 by J. J. Thomson. When he determined the mass per elementary charge of the electron, he made such thermal measurements of the kinetic energy per elementary charge in a stream of moving electrons. Then to complete the determination of the mass, he performed another experiment. He measured the motion under the influence of a known force. Although the details of that experiment were entirely different, it was equivalent to the experiment we described in Section 25 – 1, where we measured the speed of the electrons. His experiment actually measured their momentum mv, and he then found the mass m and the speed v from the known values of mv and of the kinetic energy $\frac{1}{2}mv^2$. We shall study the method he used in Chapter 28.

25–5 EMF AND THE ENERGY SUPPLIED BY A BATTERY

In the last few sections we have discussed the direction of the electric current and the measurement of its strength. Also we have seen that energy is transferred to the charged particles and may be taken from them as thermal energy. Indeed, there is no electric current without some energy transfer.

Clearly, the details of the flow of energy in a circuit are often complex. The electric field serves as a medium through which kinetic energy is supplied to the charges in the external circuit, and inside the battery charged particles are moved against the electric forces from the charges at the terminals. In the rest of this chapter, we shall look at the electric fields and at the energy in more detail. Although we shall not find it practical to establish the exact electric-field patterns for most circuits, we shall arrive at an understanding of the overall flow of energy in all sorts of circuits.

So far in this chapter we have employed batteries as the source of the energy dissipated as thermal energy elsewhere in the circuit. How much energy does a battery supply and on what does that amount depend? We shall find that a battery which uses stored chemical energy to supply electrical energy usually supplies the same energy to each elementary charge.

Consider, for example, the energy supplied by a battery to each elementary charge in an ion beam accelerated from one charged metal plate to another parallel charged plate. In Chapter 24 we learned that the electric force field between the plates is uniform, and if the plates remain connected to a battery, the force is inversely proportional to the distance between the plates. Consequently, the force on an elementary charge times the distance between the plates does not depend on the plate separation. As we have just verified in Section 25 – 4, this product Fd gives the energy transferred to an elementary charge. Hence, no matter how we change the plate separation, the energy supplied to each elementary charge by a given battery is always the same.

Not only can we change the separation between the plates, but we can also change the plate area. The electric force on an elementary charge between the plates does not change; and, as we expect, the energy supplied to each ion that passes between the plates stays the same. The battery moves more charges when it initially charges up the plates, but this does not influence the energy supplied to each elementary charge moving across in the ion beam.

The energy supplied per elementary charge does not depend on the magnitude of the current. By using a bigger source of hydrogen ions or by boiling more electrons out of a hot wire filament, we can increase the ion currents that we used in our earlier experiments. The energy per elementary charge supplied by the battery does not change. (Later we shall see more evidence that the energy per elementary charge does not depend on the strength of the current.)

We have used the ion currents because they are easy to understand. They are not essential to our result. If we connect a current meter and a high-resistance wire in series between positive and negative terminals of a battery, we can measure the number of elementary charges transported past any point in the circuit in a specified time. In such a circuit almost all the energy that the moving charges acquire from the electric field is turned immediately into thermal energy in the high-resistance wire. By measuring the heat dissipated in the wire, we can calculate the energy supplied by the battery per elementary charge (Fig. 25 – 12). We find that the energy supplied per elementary charge is the same in this different circuit as it is for an ion beam between parallel metal plates, as long as we use the same battery each time.

All these experiments indicate that a battery is a device which supplies a definite amount of energy per elementary charge; for example, as we calculated in Section 25 – 1, a 90-volt battery supplies 1.4×10^{-17} joule per elementary charge. The energy supplied per elementary charge is known as the EMF of the battery. You should read this name as "ee em eff." The old name from which the initials come is the electromotive force, but this name is misleading for we are not dealing with a force in the usual sense at all. Almost everyone now calls the energy per elementary charge the EMF rather than using the old words. We shall symbolize the EMF by the letter $\mathcal{E}$.

Suppose we take three identical batteries, each of which supplies an EMF of 1.4×10^{-17} joule per elementary charge. If we connect them in series in a circuit, each should supply 1.4×10^{-17} joule for each elementary charge. The total EMF—that is, the total amount of energy supplied by all the batteries per elementary charge—should therefore be

$$(3)(1.4 \times 10^{-17} \text{ joule/elem. ch.}) = 4.2 \times 10^{-17} \text{ joule/elem. ch.}$$

This is exactly what we find; and indeed it is exactly what you would expect on the basis of the experiments with the batteries in series that we mentioned in Section 24 - 6. There, when we connected three hundred 90-volt "B" batteries in series to make a large electric balance, we increased the force between the parallel metal plates in proportion to the number of batteries. Consequently, the batteries also supplied energy per elementary charge in proportion to their number.

Now we are in a position to answer the question of how much energy a battery supplies to an electrical circuit. This energy is just the product of $\mathcal{E}$, the EMF of the battery, times the total charge q that is transported:

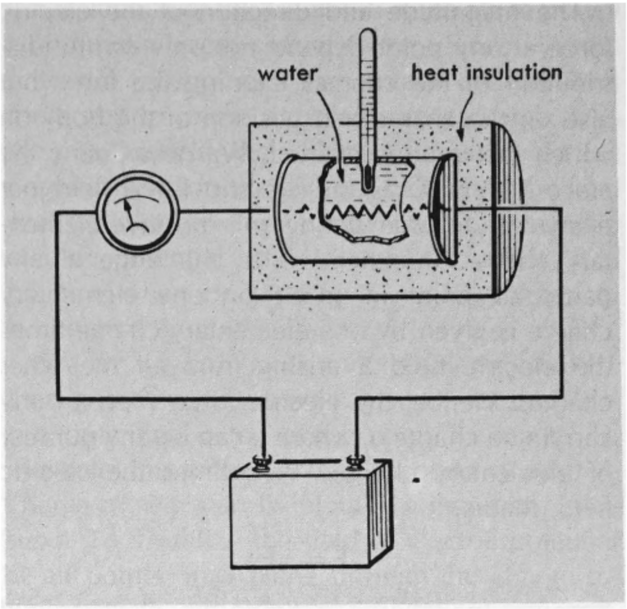
$$q\mathcal{E} = (\text{elem. ch.})(\text{joules/elem. ch.}) = \text{joules}$$

If no charges are transported the battery supplies no energy.

If instead of the total charge q we know the current I, or the rate at which charge is transported, then the rate at which the battery supplies energy is given by $I\mathcal{E}$:

$$I\mathcal{E} = (\text{elem.ch./sec})(\text{joules/elem. ch.}) = \text{joules/sec}$$

The rate at which energy is supplied to a circuit is known as the *power* supplied. The unit of power 1 joule/sec is called a watt after James Watt, the steam-engine maker. During a time interval t, the battery supplies a total energy of $I\mathcal{E}t$; in other words, the energy supplied by the battery is equal to the product of the power and the time.



25–12 We can measure the energy dissipated as heat in a resistance wire by the temperature rise of the water surrounding the wire. Since the current meter indicates the number of elementary charges that pass per second, we can calculate the energy supplied per elementary charge by the battery.

25–6 ELECTRIC FIELD AND ELECTRIC POTENTIAL

We have used batteries as the ultimate sources of stored energy. That energy is used to concentrate charges, and the electric force fields originating from these concentrations of charge in turn convey the energy to the ions moving between the metal plates or to the electrons moving in metal wires. The electric force fields are set up whenever we concentrate charges, whether we make these concentrations by the action of a battery, by rubbing, or by any other means. We should, therefore, discuss the energy of an elementary charge in an electric force field without restricting the discussions to a particular source of EMF. For this reason we shall now return to the electric force field.

In Section 24 - 3, we discussed the electric force field produced by a fixed distribution of charges. The electric force field is the pattern of electric force that would be exerted by a given distribution of charges on another charged object. In imagination we place this other charge successively at each point in space, and we either measure or compute the force that would be exerted on it at all the different points.

The magnitude and direction of the electric force at any point depend not only on the distribution of the charges exerting the force but also on the charge and position of the body on which the force is exerted. We now define the *electric field* $\vec{\mathbf{E}}$ as the electric force field per elementary charge, acting on a positive elementary charge. In other words, the force at any particular point in space on one elementary charge is given by one elementary charge times the electric field $\vec{\mathbf{E}}$ arising from all the other charges. Hence, the electric force $\vec{\mathbf{F}}$ on a body carrying a charge q, where q can be any number of elementary charges, is q times the electric field; that is,

$$\vec{\mathbf{F}} = q\vec{\mathbf{E}}$$

Because the charge q times the electric field gives the force $\vec{\mathbf{F}}$ exerted on that charge, the units of electric field are newtons per elementary charge.

From the vector equation $\vec{\mathbf{F}} = q\vec{\mathbf{E}}$ we can see that the electric field is in the same direction as the force on a positive charge, and in the opposite direction from the force on a negative charge. We can picture the electric field with the same lines we drew in Chapter 24 to picture the electric force field exerted by charged bodies, since we drew the force field lines to show the direction of the force exerted on a positive charge. For example, the electric field due to a positive point charge points outward, and the electric field due to a negative point charge points inward.

Suppose we know that a micro-microbalance exerts a force of 3.0×10^{-14} newton on each elementary charge between its plates. We can say this now in a different way: the electric field between the plates of the microbalance is 3.0×10^{-14} newton per elementary charge (pointing from the positively charged plate to the negatively charged plate). If a plastic sphere having a charge of two elementary charges is between the plates of this micro-microbalance, the electric force on it is 6.0×10^{-14} newton, but the electric field of the micro-microbalance is still 3.0×10^{-14} newton/elem. ch. The electric field depends only on the micro-microbalance and not on the charge of the plastic sphere on which the force is exerted.

The electric field of a micro-microbalance is due to the net effect of all the charges on the plates of the apparatus. Let us consider a much simpler example: the electric field around a body of charge q_1 located at a fixed point. Coulomb's law tells us that a body of charge q_2, located a distance r away from the first body, will experience a force of magnitude

$$F = \frac{kq_1q_2}{r^2} = Eq_2$$

so the electric field due to the first body has magnitude

$$E = \frac{kq_1}{r^2}$$

The direction of $\vec{\mathbf{E}}$ points outward from the first body if the charge on it is positive, inward if it is negative. Notice that the magnitude of $\vec{\mathbf{E}}$ varies from point to point in this case, while it was the same at all points between the plates of the micro-microbalance (since the force on a given charge was the same at all points between the plates). In general, the value of the electric field at a point due to a charged body depends upon where the point is in relation to the body.

As you can see, the electric field is an aid that allows us to say what the force will be on any charge placed at any position. When the field is due to a collection of charges located at one point, the equation for the field is simply a restatement of Coulomb's law. Even when the electric field is produced by charges located at many different points, the vector sum of the Coulomb forces exerted by all the charges on another charge q is equal to $q\vec{\mathbf{E}}$.

Although the electric field does not contain any new information, it is a great convenience. At times we can evaluate it without going through the detailed computation of the force based on knowing the magnitude and position of the charges. Later we shall find it even more important. For example, in Chapter 29 we shall discover electric fields that arise in other ways.

We have introduced the electric field here mainly for the purpose of discussing the energy of an elementary charge as it moves from one point in space to another. For the simple case of an elementary charge moving from one point to another in the vicinity of the charge q_1, we already know a great deal about the energy. We know about this energy because the Coulomb field of force has exactly the same form as the gravitational field of force that we studied earlier. In Section 18 – 5 we found that two masses which attract each other with a gravitational force

$$F = G\,\frac{m_1m_2}{r^2},\ \text{attractive}$$

have a corresponding gravitational potential energy

$$U_g = -\,G\,\frac{m_1m_2}{r}$$

The zero of potential energy is chosen to be at infinite separation, and the minus sign indicates that the potential energy is then negative at any separation r. The potential energy is negative because the masses attract each other as indicated.

Now we know that two charged bodies interact with an electric force

$$F = k\,\frac{q_1q_2}{r^2},\ \text{repulsive (for like charges)}$$

So in exactly the same way, they have an electric potential energy

$$U_e = k\,\frac{q_1q_2}{r}$$

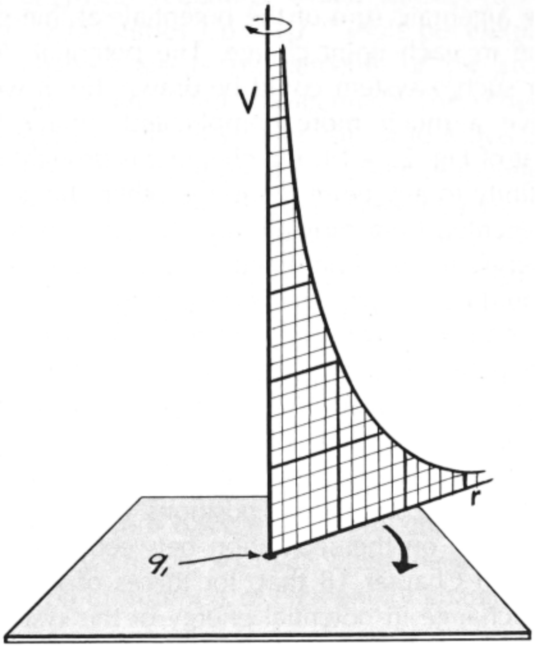
There is no minus sign in this expression because charges of the same sign repel, instead of attracting, each other like gravitating masses. The potential energy of two like charges, which has been taken as zero at infinite distance, will have a positive value for any finite value of r. The potential energy of two unlike charges is also taken to be zero at infinite separation, but is negative for finite values of r, because the charges attract.

Now, since the potential energy when the charge q_2 is a distance r from the charge q_1 is kq_1q_2/r, then the potential energy per elementary charge due to the electric field of q_1 is

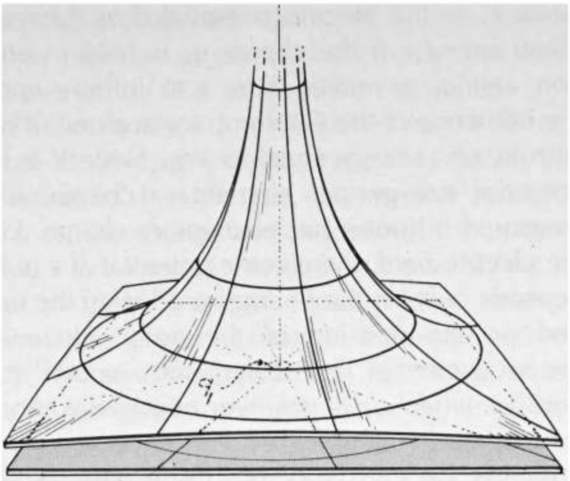
$$V = \frac{kq_1}{r}$$

We shall call this potential energy per elementary charge the *electric potential V* at the distance r, or the electric potential drop between r and infinity. If the charge q_1 is held in position and q_2 is moved from r to infinity under the influence of the Coulomb force alone, it will gain kinetic energy equal to Vq_2. Since V is the potential energy per elementary charge, it is measured in joules per elementary charge. Like the electric field, the electric potential at a point depends only on the charges producing the field and on the position of the point. However, electric potential is a scalar, not a vector. It is positive when q_1 is positive, negative when q_1 is negative.

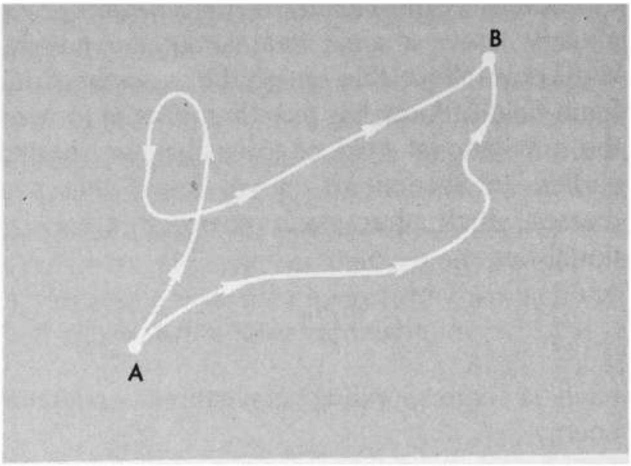
Just as we can work out the electric field $\vec{\mathbf{E}}$ throughout space, so we can work out the electric potential V for every point in space. To illustrate this potential field around a positive charge q_1, we start by plotting a graph of V versus r. To visualize the field of electric potential at all points on a plane through the charge q_1, we swing the graph around a vertical axis passing through $r = 0$ (Fig. 25 – 13). In other words, we swing it around the position of the charge q_1. As it sweeps around, the graph generates a potential "hill" whose height above any point on the plane represents the potential at that



25–13 A graph of electric potential V versus the distance r from a charge q_1 is rotated around a vertical axis passing through r = 0 in order to create the potential "hill" of Fig. 25–14.



25–14 The height of the potential "hill" directly above any point on the plane represents the electric potential at that point on the plane.



25–15 When a charge moves from a point A to a point B in an electric field produced by nailed-down charges, the change in potential energy does not depend on what path the charge takes.

point. The potential "hill" for a plane passing through a point charge is shown in Fig. 25 – 14. We shall find this potential "hill" of great value in Chapter 26 when we discuss the motion of one charged particle in the Coulomb field of another.

We can extend the idea of electric potential to include potentials which are due to more than a single point charge. At any point, the potential V of a system of several point charges is just the algebraic sum of the potentials at that point due to each point charge. The potential "hill" for such a system could be drawn, but it would have a much more complicated surface than that of Fig. 25 – 14. If a charge q is brought from infinity to any point P and the other charges are prevented from moving, the change in potential energy of the combined system, including q, would be Vq, where V is the potential at point P.

Sometimes we are interested in knowing how the potential energy of a system changes as a charge is moved from one point to another within the system. Now, the Coulomb force between each pair of charged bodies depends not on the individual positions of the bodies, but only on the separation between them. We saw in Chapter 18 that, for forces of this type, the change in potential energy of the system as one of the bodies moves along any path from some point A to another point B depends only on the positions of A and B, and not on the particular path (Fig. 25 – 15). Suppose we nail down all of the charges in the system but one, of size q, and let it move from A to B. Then the change in electric potential energy, equal to the potential energy when q is at B minus the potential energy when q is at A, is just q times the difference between the electric potentials at the points B and A due to all of the other charges. Thus,

$$\Delta U = q(V_B - V_A)$$

The quantity $(V_B - V_A)$ is called the *electric potential difference* between points B and A in the field of all the nailed-down charges.

For example, suppose that in a Millikan experiment a positive elementary charge gains a kinetic energy of 5.0×10^{-17} joule as it is accelerated from the positive plate to the negative plate. This tells us that the electric potential difference between the plates is 5.0×10^{-17} joule/elem.ch. Since the charge has gained kinetic energy and the system has lost potential energy, the potential at the negative plate is lower than the potential at the positive plate. The potential falls from the positive plate to the negative plate. If the charge were to move from the negative to the positive plate, on the other hand, it would be moving to a region of higher potential. A positive charge acted upon by the Coulomb force alone would decelerate and lose kinetic energy, while the potential energy of the system would increase.

25-7 BATTERIES, VOLTS, AND AMPERES

In Section 25 – 5 we looked at some of the evidence that batteries supply a definite amount of energy per elementary charge, but we did not examine why or how the battery maintains a constant EMF. Although we shall not attempt a detailed explanation here, we shall try to get a crude idea of how this comes about.

Batteries are made by connecting a large number of electrochemical cells in series. In a single electrochemical cell charges are forced onto the terminals against the electric repulsion of the charges already present. Their electric potential energy is thus increased. This increased potential energy of charges at the terminals is supplied by chemical energy released in chemical reactions between the constituents of the cell. In each individual chemical reaction the chemical potential energy is reduced a definite amount. Also because each individual chemical reaction involves a definite number of atoms or ions moving across the cell, this particular amount of energy is used to push a definite number of elementary charges up the electrical hills at the terminals. Therefore the cell gives a fixed energy per elementary charge. Also, a battery of cells in series gives an energy per elementary charge in direct proportion to the number of cells.

As long as the chemical composition and the positions of the constituents of a cell stay substantially the same, each individual chemical reaction is just like every other one. So a fixed EMF is maintained. When too much charge has been transferred, however, too many chemical reactions have taken place, and the composition of the cell is changed.

Then the cell no longer provides the same chemical reactions with the same ions moving through the same distances. The EMF falls off and we say that the battery has run down. In some cells, by forcing a current to pass in the opposite direction, we can reverse the chemical reactions, thus "recharging" the battery. Here "recharging" means restoring the chemical energy. We do not change the electrical charge of the battery.

The reactions that go on in electrochemical cells of many different kinds transfer about the same number of joules per individual chemical reaction (the same number of joules per elementary charge), in many cases about 1 or 2×10^{-19} joule per elementary charge. It is therefore convenient to use a practical unit of EMF of about this size. The practical unit that is commonly used is called the *volt*. We define the volt as 1.60×10^{-19} joule per elementary charge. A flashlight dry cell, for example, gives 1.5 volts or 2.4×10^{-19} joule per elementary charge; and, as we know, the 90-volt balancing battery gives $90 \times 1.6 \times 10^{-19} = 1.4 \times 10^{-17}$ joule per elementary charge.*

We have defined the volt as a certain number of joules per elementary charge. Thus if we want to measure the number of volts supplied by a given battery, we must perform an experiment using this battery to provide energy to a measured number of charges. However, we need to do this only once, because we can keep our calibrated battery as a standard, just as we can keep a good meter stick on hand to use in distance measurements. (Prescriptions for making standard cells, the Weston cell for instance, will be found in treatises on electricity.)

Using the volt at the unit of EMF will prove convenient not only in practical work dealing with electrochemical cells or batteries, but also in studying the fundamental structure of atoms. The chemical reactions which transfer a few volts—a few times 1.6×10^{-19} joule per elementary charge—are rearrangements of the structure of atoms and molecules. The energy required to pull out an electron or to move one around from one atom to another is therefore about a volt times an elementary charge. For this reason, physicists often state energies in electron volts. These are parcels of 1.6×10^{-19} joule, and they are of just the right size to measure the energies of individual chemical reactions or the energies required to ionize individual atoms. In the box on page 480 we have estimated the energy needed to ionize an atom, and it is a reasonable number of electron volts.

We chose our basic unit of EMF, the joule per elementary charge, so that, when the charge is measured in numbers of elementary charges, $q\mathcal{E}$

---

*We have defined the volt as 1.60×10^{-19} joule per elementary charge. Often the practical unit of current is defined first; and the volt is defined in terms of it. As we shall see later, the two definitions agree.

gives the energy in joules. We would like to continue to measure the energy in joules even when we shift the unit of EMF to the practical unit, the volt (or 1.60×10^{-19} joule per elementary charge). We therefore introduce a new unit of charge. We design the new unit, the coulomb, so that one coulomb times an EMF of one volt gives one joule,

$$(1 \text{ coul})(1 \text{ volt}) = 1 \text{ joule}$$

Since 1 volt is 1.60×10^{-19} joule per elementary charge, this is

$$(1 \text{ coul})(1.60 \times 10^{-19} \text{ joule/elem. ch.}) = 1 \text{ joule}$$

Consequently,

$$1 \text{ coulomb} = \frac{1 \text{ joule}}{1.60 \times 10^{-19} \text{ joule/elem. ch.}} = 6.25 \times 10^{18} \text{ elem. ch.}$$

The unit of current that goes along with this unit of charge is called an *ampere*. An ampere is a flow of charge of 1 coulomb per second. Therefore 1 ampere = 6.25×10^{18} elementary charges per second.

When we deal with light bulbs or electric irons the ampere is a more convenient unit of current than the basic current unit of one elementary charge per second. The current that runs an ordinary light bulb is about an ampere. It would be clumsy to say that the lamp current is about 10^{19} elementary charges per second. Most current meters are calibrated in amperes, that is, in units of 6.25×10^{18} elementary charges per second. The common name for a current meter is an ammeter.

The volt and the ampere are designed to fit together in the same way that the joule per elementary charge fits with a current measured in elementary charges per second.

In other words, when the current is measured in elementary charges per second and the EMF in joules per elementary charge the power is $I\mathcal{E}$ watts. In just the same way, if we measure the current in amperes (coulombs per second) and the EMF in volts (joules per coulomb), the power $I\mathcal{E}$ again is given in watts. The practical units fit together in the same way as the natural units, because we have designed the coulomb so that, when a charge of 1 coulomb moves through an EMF of 1 volt, a joule of energy is transferred.

Electron Volts and Atoms

We can see that the volt (1.6×10^{-19} *joule per elementary charge*) and the electron volt (1.6×10^{-19} *joule*) are of the right size for atomic affairs by computing the electric potential energy of an electron at the outside of an atom. Since atoms are neutral, such an electron must be acted on by the electric force from one positive elementary charge at about the distance given by the atomic radius. For hydrogen it is the force exerted by the proton. We know that atomic dimensions are in the range of an angstrom, that is 10^{-10} meter. Then since the charges are of opposite sign and each is one elementary charge in magnitude, the electric potential energy is

$$U_e = k\,\frac{q_1 q_2}{r}$$

$$= \frac{\left(2.3 \times 10^{-28} \dfrac{\text{newton-m}^2}{\text{elem. ch.})^2}\right)(1 \text{ elem. ch.})(-1 \text{ elem. ch.})}{10^{-10} \text{ m}}$$

$$= -2.3 \times 10^{-18} \text{ joule}$$

This energy may not be the only energy. For instance, in a planetary model, as we know from Chapter 18, the kinetic energy of the electron is half as great and, of course, positive. But in any case this electric potential energy gives us an estimate of the order of magnitude of the energy we must supply in order to take the electron away from the atom. Therefore, the binding or ionization energy of an atom is of the order of 10^{-18} joule. Since electron volts are 1.6×10^{-19} joule at a throw, this estimate is about 6 electron volts. The actual amount of energy necessary to ionize an atom depends on the kind of atom, but it runs from about 2 to about 20 electron volts. This indicates that the electric forces holding the atom together account for the energy necessary to rearrange or pull apart the electric particles of which it is made.

Motors, light bulbs, and other electrical appliances are usually designed to use energy at a particular rate. The power, which is the energy they consume per unit time, is usually specified; and so is the voltage at which they will use this power. From the power and the voltage you can determine what current they will draw. For example, a 60-watt, 110-volt light bulb will draw a current of $\frac{60}{110} = 0.545$ ampere when an EMF of 110 volts is placed across its terminals.

Although the volt is a convenient unit of EMF and the ampere is a convenient unit of current, the coulomb turns out to be an appallingly large unit of charge. If two charges, each one a coulomb, are placed a meter apart, the force between them is about ten billion newtons. When dealing with the forces between charges, therefore, it is often more convenient to use the microcoulomb (10^{-6} coulomb) as the unit of charge. A toy balloon charged by being rubbed on a woolly sleeve might have a charge of a few tenths of a microcoulomb.

25–8 STATUS REPORT

In this chapter we have considered charges that move and energy transferred under the influence of the Coulomb forces exerted by charges at rest. In Section 25 – 4, we found that as long as the charges exerting the force were at rest, the electric force on a moving charge is the same as the force on a charge standing still at the same position. Because the Coulomb forces depend only on the separations between the charges, there is a well-defined electrical potential energy.

Except for the forces exerted by batteries or generators, in this chapter we have used only the same electric forces that we had studied in the last chapter. Now in considering the forces, two problems remain. In the first place, what can we say about the forces on charges moving through batteries and generators, the forces that apparently move the charges against the Coulomb forces exerted by the concentrations of charge standing still? Secondly, what happens when a charge exerting a force on another charge is itself in motion?

In Chapter 28 we shall see that Coulomb's law does not adequately describe the force between two charged particles when both of the particles are moving. Then we find that in addition to the Coulomb force, a new force appears—a force that depends upon the motions. This additional force is of great practical importance, for it runs most of our electrical machinery. Furthermore, as we shall see in Chapter 29, this same additional force is responsible for the EMF of electric generators. Finally we are left with the forces and the EMF's supplied by batteries. As we have indicated here, their explanation can be given only in terms of the energy transfers taking place in chemical reactions. In other words, the energy supplied by a battery can be understood only when we look inside the atoms. We shall just be able to start examining the internal energy of atoms in the last chapter of the book.

In Chapter 27 we shall look in more detail at energy flow and electric currents. We shall be especially concerned with the application of these ideas to electric circuits.

QUESTIONS

1.* Suppose we use two 90-volt batteries in the apparatus of Fig. 25 – 1, and use doubly charged ions; what energy will each ion have when it hits the negative plate?

2. Two large parallel metal plates, 6.2 cm apart in a vacuum tube, are connected to the terminals of two 90-volt batteries in series. A doubly charged oxygen ion starts from rest at the surface of one plate and is accelerated across to the other plate.
   (a) With what kinetic energy $\frac{1}{2}mv^2$ does the oxygen ion crash into the other plate?
   (b) If the ion starts halfway between the plates, with what kinetic energy does it hit the negative plate?

3.* If singly charged helium ions were put through the apparatus of Fig. 25 – 4, what would be the time of flight?

4. In a time-of-flight experiment like that described in this chapter, you have adjusted the sweep circuit of the oscilloscope so that the beam moves across the tube in 0.5 sec. The total length of the trace made in that time is 10 cm. The contact points on the track for the rolling ball are 1.2 m apart. If the spikes on the oscilloscope trace are 8.0 cm apart, what is the speed of the ball?

5. In the apparatus used to measure the mass of the positive hydrogen ion (proton) of Section 25 – 1, you use a source of heavy hydrogen and observe a spike which corresponds to ions arriving with a time of flight 1.4 times that of the usual positive hydrogen ions.

   (a) What is their mass if they are singly charged?

   (b) If doubly charged?

   (c) Could they be singly charged helium ions? Doubly charged helium ions?

   (The mass of a helium atom is four times that of a hydrogen atom.)

6.* If 2 × 10^{16} pairs of singly charged ions are produced per second and reach the plates of Fig. 25 – 10 without recombining, what is the current in the battery connected to the plates?

7.* A poorly made diode tube has considerable residual gas left inside. When the filament is heated and a battery connected between filament and anode, there is a current within the tube consisting of both electrons and positive ions. Consider a plane close to the filament. In each second, 3.0 × 10^{16} electrons are emitted by the filament and move across the plane toward the anode. Also, 0.6 × 10^{16} singly charged positive ions per second pass through this surface in the opposite direction. What is the current through the plane?

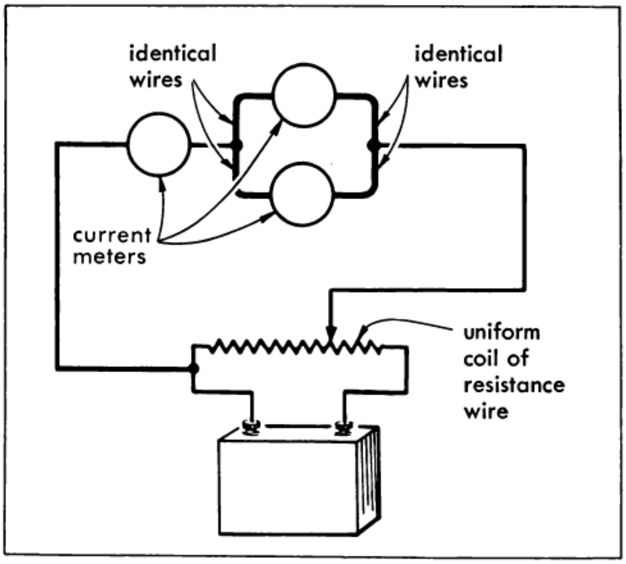
8. You ionize some gas between parallel metal plates, making equal numbers of negative and positive ions in every small volume between the plates. There is a battery and a current meter connected in series with the plates, and a steady current is flowing through the circuit. The positive plate is on the left, and the negative on the right.

   (a) In what direction do you expect the electric current to go between the plates?

   (b) If the current meter measures 10^{16} elementary charges per second, how many positive elementary charges are brought per second by the ions coming to the negative plate?

   (c) How many negative elementary charges are coming per second to the negative plate?

   (d) In one second, how many positive ions cross to the right through a plane halfway between the plates?



25–16 **For Question 10.**

9. You are given some metal wire, batteries, and a calibrated current meter. How can you calibrate a second current meter?

10. You are given a half dozen identical current meters, one of which has a single calibration point—that is, a mark correctly indicating when a 1-ampere current (a current of 6.25 × 10^{18} elementary charges per second) is passing through it. You also have a few batteries and a large number of identical metal wires. Answer the following questions to complete the calibration of the current meter.

    (a) Assume you have marked a number of current meters to read 1 ampere. Then you connect three as in Fig. 25 – 16. Explain how you can get a 2-ampere calibration point on one of them.

    (b) Make sketches showing how to connect the current meters in order to mark one scale at 3 and 4 amperes.

    (c) How would you connect the current meters so as to calibrate one of them to read $\frac{1}{2}$ and $\frac{1}{3}$ amperes? (Use sketches.)

11.* When 3.00 × 10^{22} atoms of aluminum are deposited in the apparatus of Fig. 25 – 11, how many atoms of copper will be deposited?

12. A current meter has been built and you want to calibrate its scale. It is connected in series with an electrolytic cell containing a solution of silver nitrate ($Ag^+NO_3^-$) and a battery. A steady current which deflects the meter to the upper part of the scale flows for half an hour. At the end of this period, it is found by weighing that the negative electrode has 1.92 gm of silver deposited on it.

    (a) How many elementary electric charges per second passed through the circuit? (*Note:* A mole of silver has a mass of 108 grams.)

(b) What would you do to complete the calibration of the instrument?

13.* A doubly charged ion is accelerated from rest by a force of 3.10×10^{-15} newton over a distance of 9.3 millimeters. How much kinetic energy does the ion gain?

14.* Other measurements on the ion of the previous problem, after it had been accelerated, show its momentum to be 1.24×10^{-21} kg meters/second. What is the mass of the ion?

15. Two charged balls, *A* and *B*, are fastened to the ends of a nylon thread 6 cm long, which is stretched horizontally between them (Fig. 25–17). Ball *A* carries $+1.0 \times 10^{10}$ elementary charges, ball *B*, $+4.0 \times 10^{10}$ elementary charges. A bead *C*, which carries $+1.0 \times 10^{10}$ elementary charges, can slide without friction along the thread.

(a) At what point along the thread is the net electrical force on bead *C* equal to zero?

(b) What is the direction of the force on bead *C* when it is moved nearer *A* from its rest position? Nearer *B*? Is the situation stable or unstable when it is in the rest position?

(c) If bead *C* carried -1.0×10^{10} elementary charges; what would be your answers to parts (a) and (b)?

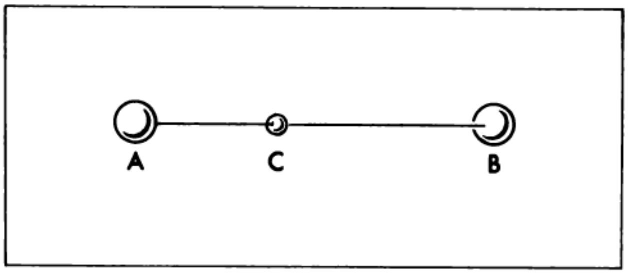
16. In Fig. 25–18, electrons emitted at *F* are accelerated toward *B*.

(a) How much energy will the electrons have which hit plate *B*?

(b) What is the electric field between *A* and *B* and between *B* and *C*?

(c) Some will travel through the hole in *B*. Where will they go?

17. With the same apparatus as in Question 2, how many doubly charged oxygen ions would have to go from the positive plate to the negative plate before that plate is heated up by one joule by the dissipation of kinetic energy of the ions?



25–17 For Question 15.

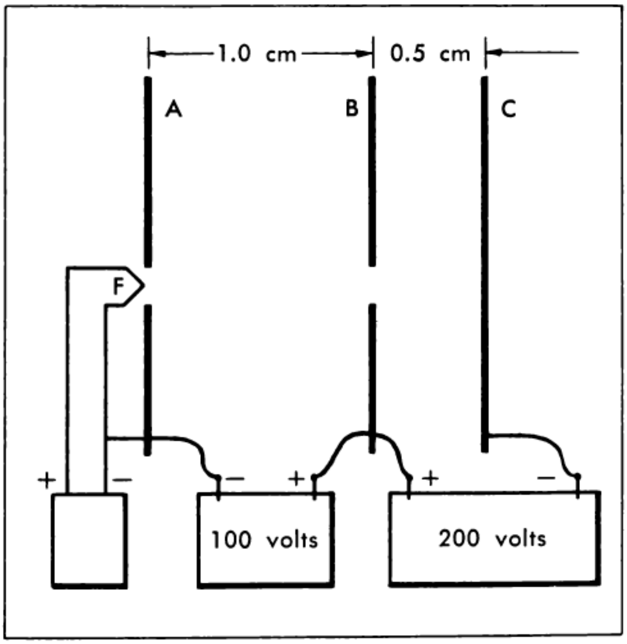
18. Suppose the doubly charged ions in the last problem were made by ripping electrons off oxygen atoms located halfway between the plates. Then the oxygen ions are accelerated in one direction and the electrons are accelerated in the opposite direction from the midpoint between the plates. How much energy will the electrons have dissipated heating the positive plate in the time that one joule is dissipated at the negative plate?

19. What evidence do you have that all elementary charges are the same? Be prepared to discuss it in class.

20.* Three 90-volt batteries are connected in series across two parallel plates 1.00 mm apart. What is the force on an elementary charge between the plates?

21.* A battery with an EMF of 2.9×10^{-17} joule/elementary charge is transporting charge at the rate of 6.0×10^{16} elementary charges/second. What is the power supplied by the battery?

22.* A single elementary charge is now at a distance *r* from a fixed charge *q*. If the elementary charge is allowed to move freely away under the repulsion of the Coulomb field, what will be its kinetic energy at a very large separation?



25–18 For Question 16.

23.* An excess of 1×10^{10} negative elementary charges is placed on a small object. What would be the electric field and electric potential at a point 0.5 m away?

24. (a) Find the electric and the gravitational forces on an electron and on a proton between two large parallel metal plates, 20 cm apart. The potential difference between the plates is maintained by 20,000 volts' worth of battery. The plates are horizontal; consider the upper plate positive.

(b) Move the particles up 1 cm. What are the forces now?

(c) What are the changes in gravitational potential energy and in electric potential energy in this motion?

25. In nuclear fission, the nucleus of an atom of uranium 238 sometimes breaks into just two fragments, owing to the Coulomb force of repulsion between its protons. Imagine that the split has just happened, and the two fragments are still touching each other (roughly 10^{-14} m apart, center-to-center) and at rest. Find their total kinetic energy when they are very far apart, if one has $Z_1 = 42$ elementary charges and the other has $Z_2 = 50$ elementary charges. Give your answer in joules and in millions of electron volts (Mev).

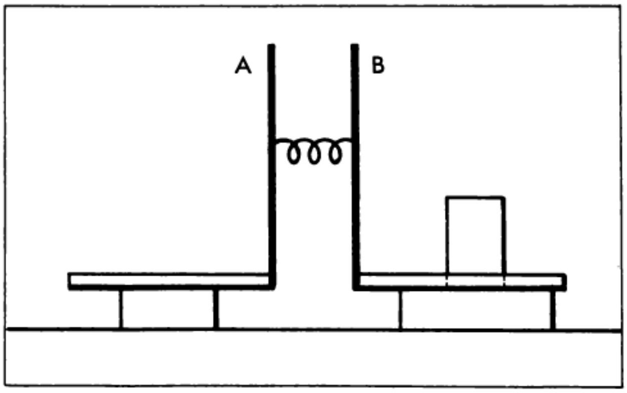
26. (a) Two protons are released from rest 10^{-10} m apart. What will their kinetic energy be at large separation?

(b) Suppose one of the protons is nailed down. What will the kinetic energy be at large separation?

27. Two vertical metal plates stand parallel to each other (Fig. 25 – 19). Both are mounted on insulating bases, but plate *A* is bolted to the table and plate *B* is set on a Dry Ice puck so that it can move freely. They are connected by an insulating spring which can be either stretched or compressed.

(a) Plate *B* is held fast in such a position that there is no force exerted by the spring, and the two plates are charged by connecting them to a battery. The battery is disconnected and plate *B* is let go. Describe the motion qualitatively but in detail.

(b) Can you think of a purely mechanical system which behaves in the same way?



25–19 For Question 27.

28. Two identical charges q are nailed in place on the y axis, equal distances a above and below the origin of coordinates.

(a) Find an expression for the potential energy due to the interaction of an elementary charge on the x axis, a distance x from the origin, with one of the nailed-down charges; with both of the nailed-down charges.

(*Note:* Measure the potential energy with the zero chosen when the elementary charge is at infinity. Express your answer in terms of the constant k in Coulomb's law, the distances a and x, and the charge q.)

(b) What would be the potential energy due to the interaction of the elementary charge if the two nailed-down charges were equal and opposite?

29. In the Millikan experiment, a latex sphere 1.3 microns in diameter was suspended between the two plates when the difference in potential between them amounted to 175 volts. From other observations on the same sphere it was known that in this case the sphere carried three elementary charges. The separation of the charged plates was 7.0 mm and the density of the latex was 1.05 gm/cm³. Use this information to calculate how many coulombs there are in one elementary charge.

30.* A battery connected to an electric heater supplies 3.12×10^{19} elementary charges per second and produces an output of 60 joules per second (watts). What is the EMF of the battery in joules per elementary charge? In volts?

31.* What is the current in amperes when 3.12×10^{19} elementary charges per second flow in a circuit?

32.* In Question 27 of Chapter 18, we found that the binding energy of the electron to the proton was 2.3×10^{-18} joule. How many electron volts is this?

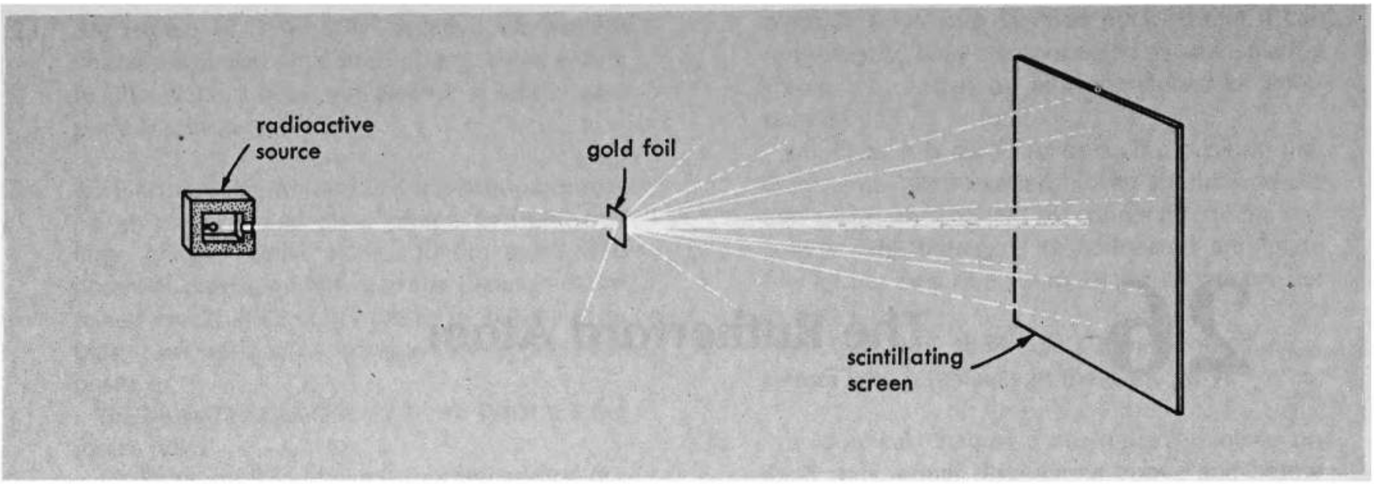
26 The Rutherford Atom

WE ALREADY know that atoms contain electrons which are far lighter than the lightest of atoms, that of hydrogen. We can rip off electrons, leaving positive ions of almost the entire atomic mass, and sometimes we can add an electron to a neutral atom, producing a negative ion. But what does an atom look like? How is it built? About these questions we have said rather little. Now that we know some things about electric energy and charged particles, we can use the latter to probe inside the atom and learn something about its structure. As we shall see, the evidence gained by bombarding atoms with charged particles to probe their depths indicates that atoms are built on about one hundred standard patterns; and in some ways these patterns are fairly simple. For this reason, with relatively few probing experiments we can learn a good deal.

In order to bring back information about the inside of an atom, our probes must get in. Consequently, other atoms and molecules, which bounce off the atomic surface, are poor probes. Also, low-energy charged particles are not of much use, for they, too, do not penetrate very far. On the other hand, a probe that passes straight through anything will tell us nothing about the place where it has been. Highly penetrating X rays or the even more penetrating gamma rays from radioactive disintegration are hard to use because of this defect. In the early years of this century, the most suitable probes for poking in atoms were alpha particles. These come out of radioactive materials with high enough energy to penetrate well into an atom; and although they often pass on through with little change of direction, they sometimes are violently deflected. Then we assume that they have hit something relatively massive inside and bounced away.

26–1 THE DEFLECTION OF ALPHA PARTICLES AND THE RUTHERFORD MODEL OF ATOMS

Figure 26 – 1 shows the setup for an experiment to probe gold atoms with high-speed alpha particles, which are emitted in the radioactive dis-



26–1 **An experiment to probe the interiors of atoms with alpha-particle "bullets."**

integration of a piece of polonium. This is our gun. Using the method outlined in Section 25 – 1, we can measure the speed of the alpha particles from polonium by measuring their time of flight along an evacuated tube. We would find that the alpha particles, which are doubly charged helium ions, have a speed of 1.6×10^7 meters per second. We could check that this is the correct value by using the apparatus described in Section 25 – 1. Suppose we direct a narrow beam of these alpha particles toward a thin gold foil and on the far side of the gold foil we place a detector of alpha particles, which will give us a signal telling us where the alpha particles go after passing through the foil.

The gold foil is a big stack of gold atoms. A suitable gold foil is so thin that each square meter has a mass of only 2×10^{-3} kg. From the density of gold we can find the number of gold atoms that are in the way of an alpha particle as it passes through the foil. The density of gold is about 2×10^4 kg/m^3. The foil thickness is therefore about

$$\frac{2 \times 10^{-3}\ \text{kg/m}^2}{2 \times 10^4\ \text{kg/m}^3} = 1 \times 10^{-7}\ \text{meter}$$

Also, we know that the mass of a gold atom is about 3.3×10^{-25} kg. In this way we learn that each atom—stacked in contact with its neighbors—has a volume of

$$\frac{3.3 \times 10^{-25}\ \text{kg}}{2 \times 10^4\ \text{kg/m}^3} = 16 \times 10^{-30}\ \text{meter}^3$$

If we think of the atoms as boxes, the side of each of the little boxes must be about 2.5×10^{-10} meter. Therefore, the thickness of the gold foil is about

$$\frac{1 \times 10^{-7}\ \text{meter}}{2.5 \times 10^{-10}\ \text{meter}} = 400 \text{ gold atoms thick}$$

The first thing we learn with this apparatus is that most of the alpha particles pass through the gold leaf without appreciable change in their direction of motion. In other words, an alpha particle usually passes right through some 400 atoms. We can conclude that most of the inside of the atom is empty of hard, massive objects from which the alpha particles would bounce off at an angle. The atom does, however, have something inside, for we see that the alpha particles gradually lose their energy in passing through gold. With any ordinary thickness of gold the alpha particles would never come out of the far side. As we can easily tell in other experiments, the alpha particle energy is lost in ionizing the gold atoms. When an alpha particle moves through the atoms, losing energy to rip off electrons, it is slowed up, much as a bullet is slowed up when it goes through many bales of hay.

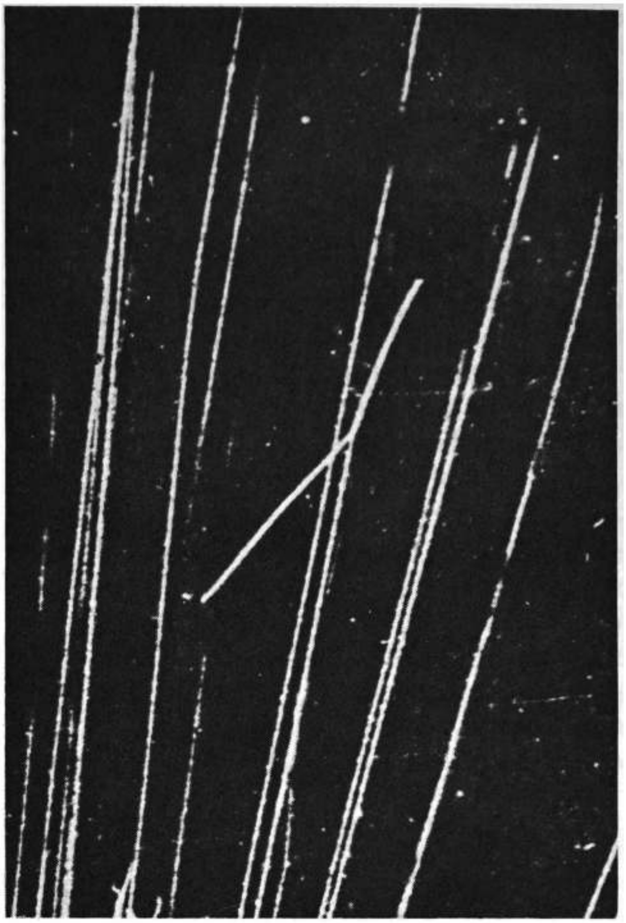
Because we have weighed atoms, however, we know that something massive must be located somewhere in the atom. We can hardly believe that this mass is made of electrons. Many thousands of electrons would be required for a light atom, and at most we can get only a few when we disrupt the atom. A gold atom has a mass about 50 times that of an alpha particle, and if an alpha particle hits such a mass head-on, it should bounce straight back. Let us, therefore, look again at the beam of alpha particles passing through the gold foil. If we look carefully all around in every direction, we find that in passing through the gold foil, about 1 out of 10,000 of the incident alpha particles is actually knocked off its original course and deflected by an angle of more than 10 degrees. Every once in a while, one out of some astronomical number of alpha particles is bounced out of its course by 90 degrees or more. The same kind of event happens once in a while when alpha particles are passing through the gas in a cloud chamber. There, if we look at myriads of cloud-chamber pictures, we eventually find a few in which the alpha particles clearly ricochet from an atom, having made almost an about-face. Figure 26 – 2 is such a photograph.

The alpha particles that are sharply deflected in the gold foil also have made just one violent collision with some part of one atom in the foil. We can be sure that it is one collision which is important, because most of the alpha particles pass straight through. Consequently, we know that the chance of even one big collision is tiny and the chance of more than one real hit on the way through is absolutely negligible.

An alpha particle which has undergone such a large deflection has been acted upon by strong forces. Somewhere inside the atom, therefore, there is at least one heavy massive center, and those alpha particles which come right at it interact strongly with it.

To account for these observations any atomic model must have two features: Because almost all the alpha particles pass right through an atom, most of the atom must be empty of anything massive. Because an occasional alpha particle bounces off backward, somewhere inside the atom there must be at least one lump of mass with which alpha particles can interact strongly. In addition, we have associated the slowing down of the alpha particles in a thick sheet of gold with the light electrons which are ripped off when alpha particles pass through atoms and ionize them. So we are tempted to associate the main mass of the atom with the positive charge which must be present in order that the atom as a whole shall normally be neutral.

In 1911 Rutherford invented an atomic model which incorporated all these features and in addition called only upon the Coulomb force to explain the interaction of alpha particles with the positive massive atomic core. His model describes an atom as a miniature solar system with a core, or *nucleus,* at the center and a number of electrons around it. The nucleus is positively charged and carries almost all of the atomic mass. The light, negatively charged electrons revolve around the nucleus, held by the Coulomb attraction, much as the planets revolve around the sun, held by gravitational attraction. Outside the atom, these negative electrons cancel the effect of the positive charge of the nucleus, so that the atom as a whole is neutral. This means that the nucleus carries a number of positive elementary charges equal to the number

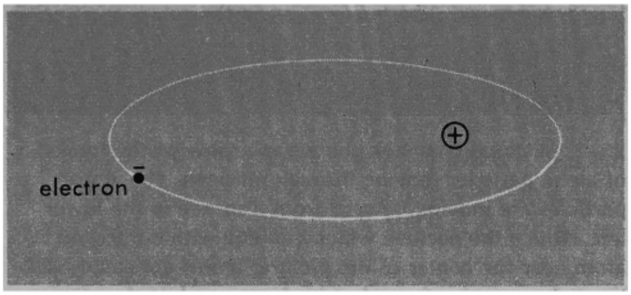


26–2 A cloud-chamber photograph showing the tracks of alpha particles moving through nitrogen. The alpha particles are moving upward from a source in the lower left. Notice the particle which collided with a nitrogen atom near the center of the picture. It was deflected back through an angle of nearly 142 degrees, while the nitrogen atom recoiled. (From P. M. S. Blackett and D. S. Lee, in Proceedings of the Royal Society, 134A, 658, 1931.)

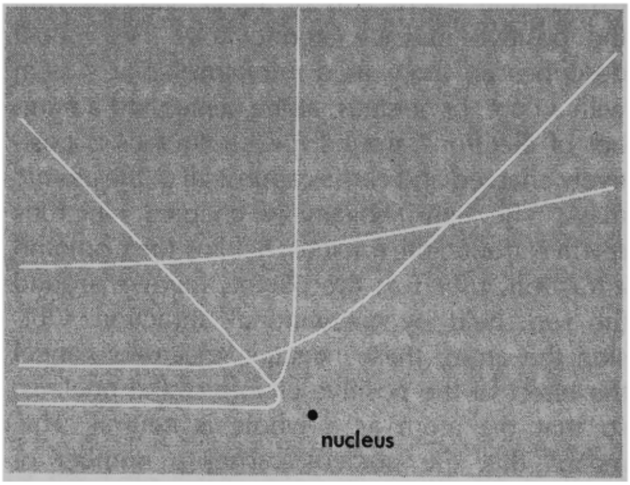
of electrons.

We have already concluded that the hydrogen atom is composed of a proton and an electron (Section 25 – 1). Thus, in the Rutherford model of the hydrogen atom (Fig. 26 – 3), the proton is the nucleus, and there is one electron moving around it in an orbit. Similarly, the helium atom consists of an alpha particle and two electrons. In helium the alpha particle is the nucleus, and two electron planets revolve about it.

In the Rutherford model of any atom the dimensions of the nucleus and of the electrons are assumed to be very small compared with the overall size of the atom, so that most of the atom's volume is actually empty space. In this space near the nucleus, the electric field is essentially that of the nuclear charge, a Coulomb's-law field whose strength varies as the inverse square of the distance from the nucleus. The planetary electrons are found in the outer regions of the atom, and there these negative charges will also contribute to the electric field. Outside the atom these negative charges completely screen the positive charge of the nucleus, and the electric field vanishes.



26–3 The Rutherford model of the hydrogen atom. On this scale both the electron and the nucleus are so very small that in fact you could not see them at all, even though the linear magnification is 10^9 times.



26–4 The computed paths of alpha particles approaching a nucleus. These were computed using an inverse-square force of repulsion.

In this model, an alpha particle fired through an atom some distance off center will pass through without appreciable deflection. It will just brush the electrons aside. But in a stream of alpha particles fired at gold leaf, a few will pass so near the center of an atom that they encounter the strong electric field close to its heavy nucleus—they will be "scattered" through a large angle.

We see that Rutherford's model accounts, at least qualitatively, for the experimental facts. It explains why atoms let most alpha particles go through almost unaffected, and why a few alpha particles are strongly deflected. Before we place more confidence in this model, however, we must submit it to a quantitative test. This Rutherford and his students did with great skill.

26–2 THE TRAJECTORIES OF ALPHA PARTICLES IN THE ELECTRIC FIELD OF A NUCLEUS

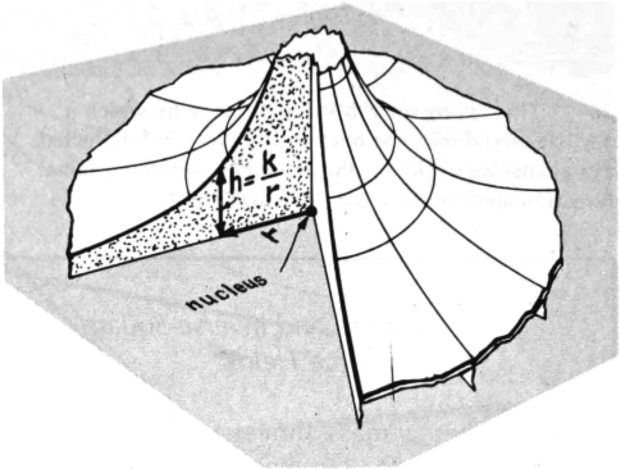
In Rutherford's model an alpha particle near a nucleus experiences a force of repulsion, inversely proportional to the square of the distance. To test this model, we must work out the paths that an alpha particle would take in such a force field.

Examples of the computed trajectories are shown in Fig. 26 – 4. Although we shall not reproduce the long computations of these paths, they involve nothing more mysterious than Coulomb's law, Newton's law of motion, and the detailed geometrical arrangement. The curves are hyperbolas, all with the nucleus at the focus. (The hyperbolic orbits here are related to the elliptic orbits of planets in the inverse-square gravitational force field. For some remarks on this relation, see the box on page 488.)

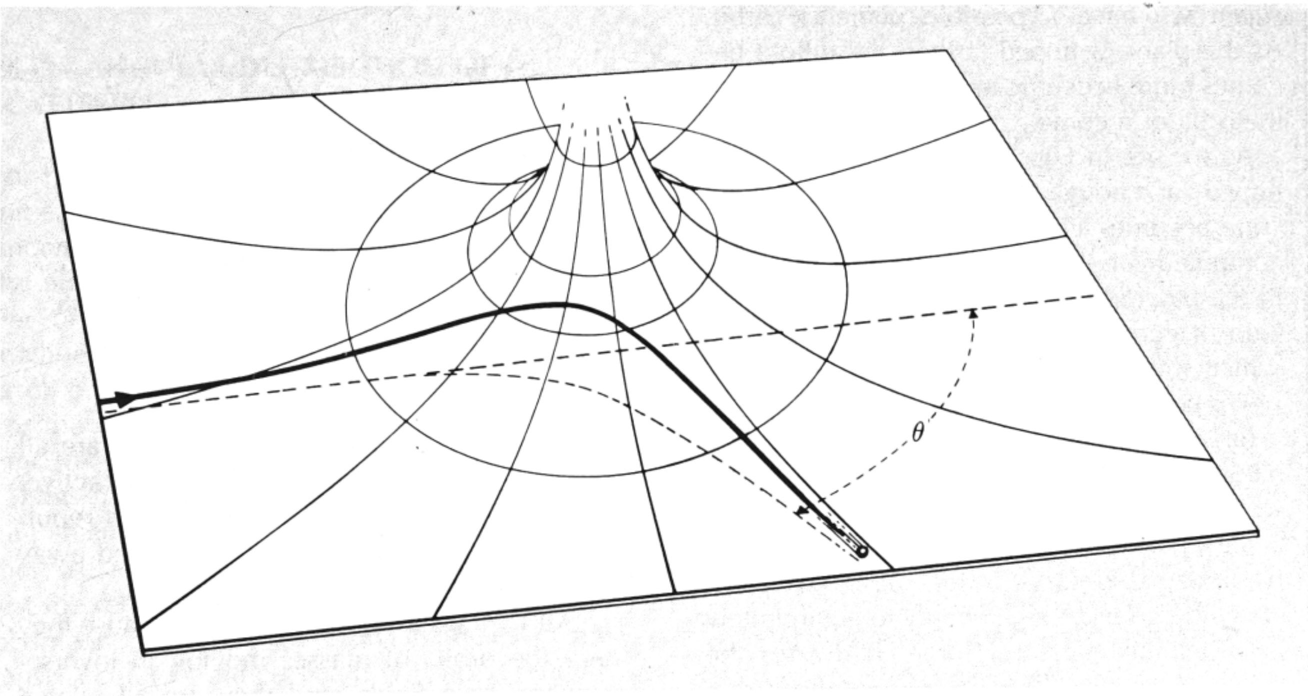
You may also convince yourself that the computed results are sensible by looking at a mechanical model which reproduces the essentials of alpha-particle scattering in the Coulomb-force field. You may even want to construct such a model and experiment with it. In the mechanical model the alpha particle is represented by a steel ball which rolls with little friction on a smooth curved hill rising from a level table (Fig. 26 – 5). We make the gravitational potential energy of the ball on the hill correspond to the electric potential energy of the alpha particle near the nucleus; so this model pictures a plane through the center of an atom with the third dimension, the height in the model, representing not space but potential energy. The electric potential energy of the alpha particle varies as $1/r$, where r is its distance from the nucleus. Therefore we build our hill so that above any point on the plane the height is inversely proportional to the distance r of that point from the center.

With this model, the motion of an alpha particle on the shaded plane through the nucleus in Fig. 26 – 5 is modeled by the motion of a ball on the potential-energy hill above. If we look down on a ball rolling on a real hill of this shape, the path we see will closely approximate that of an alpha particle near the nucleus. (See Fig. 26 – 6.)

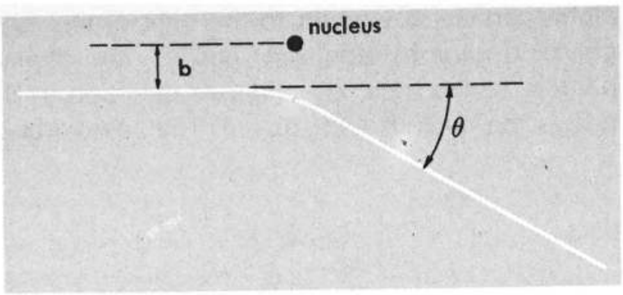
An alpha particle aimed head on at a nucleus will be scattered in exactly the backward direction. Similarly, in the mechanical model, if we start the ball rolling exactly toward the center of the hill, it will go straight up, to a height where its potential energy equals its original kinetic energy. There it reverses its motion and returns to the starting point. If, however, we aim the alpha particle at a point to the right of the nucleus, Coulomb repulsion pushes the alpha particle to the right, changing its direction as it passes the nucleus (Fig. 26 – 6). The angle θ be-



26–5 A mechanical model to illustrate the paths of alpha particles near a nucleus. Alpha particles have a potential energy proportional to $1/r$, where r is the distance from the nucleus. The hill is so constructed that the height of any point on its surface is proportional to $1/r$ where r is the distance on the plane from the center. Consequently, a ball on the hill will have gravitational potential energy proportional to $1/r$. Its motion then resembles the motion of a charge moving on the plane in the electric field of the nucleus.



26–6 The solid line shows a possible path of a rolling ball in the "hill" model of the Coulomb field. The dashed straight line on the plane below the hill shows the direction in which one ball was originally aimed. If you look straight down on the hill, the apparent path of a ball models the path of an alpha particle moving in the horizontal plane. This path is the dashed curved line on the plane.



26–7 **The aiming error b is the distance by which a particle would miss the nucleus if it were not deflected. The scattering angle θ is the angle between the original direction and the direction after deflection.**

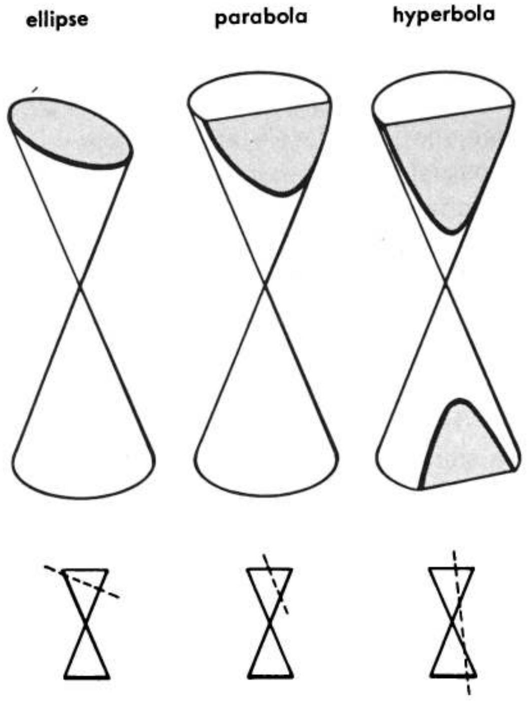
tween the final direction of motion and the original line of flight is called the scattering angle (Fig. 26 – 7). The closer the alpha particle comes to the nucleus, the stronger the Coulomb force it feels and the more nearly that force pushes straight backward against the original direction of motion. Therefore, the scattering angle θ will *increase* when the "aiming error," represented by the distance *b*, *decreases*. Figure 26 – 9 shows the corresponding motions of the ball in our mechanical model.

Conic Sections and Inverse-Square Force Fields

Whenever a mass moves under the influence of a force that varies as the inverse square of the separation from a fixed center, the orbit is a conic section: it is a curve which can be made by a plane slicing through the surface of a cone. If the plane slices through the cone perpendicular to its axis, the surface of the cone is cut in a circle. As you know, this orbit is a possible planetary orbit. If the plane is now tipped slightly, the surface of the cone is sliced in an ellipse (Fig. 26 – 8). Again we have a possible planetary orbit. As the plane is tipped farther, the ellipse becomes more eccentric and the orbit resembles the orbit of a comet.

As we see in Fig. 26 – 8 when the plane is tipped far enough, the intersection with the cone becomes a hyperbola, a curve like those computed for the orbit of an alpha particle in the electric field of the nucleus. The same kind of curve also gives the orbit of a mass which moves past the sun so fast that it will never come back.

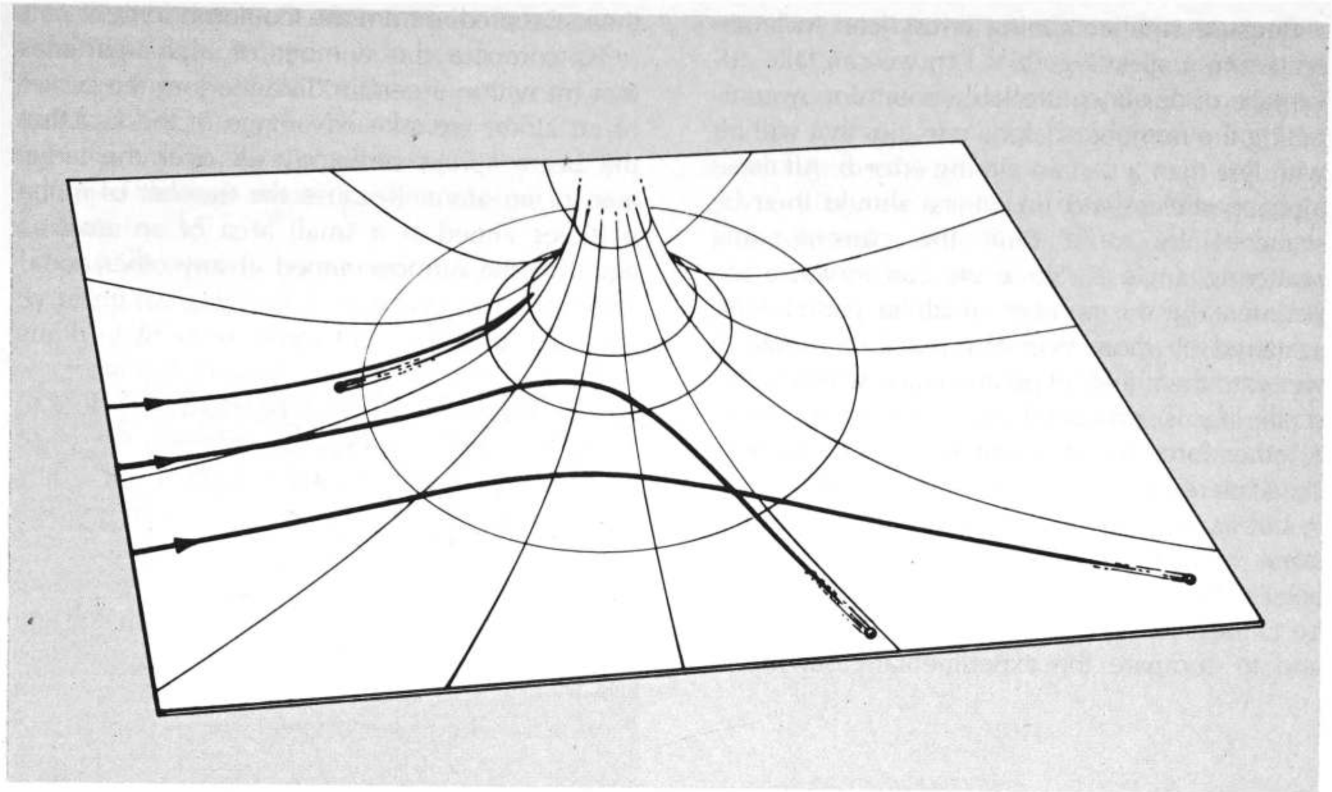
Just in between the closed orbits which are ellipses and the open hyperbolic orbits on which the mass never returns, there is one angle at which the plane cuts the cone in a parabola. This curve is the orbit of a particle with barely enough energy to get infinitely far from the center of force. It divides the orbits of masses that return from those that make a single trip.



26–8 **The conic sections. Plane figures such as the ellipse, the parabola, and the hyperbola are formed by the intersections of planes at various angles with a cone.**

Ellipses, parabolas, and hyperbolas are all possible orbits when the force is attractive. Only hyperbolas will be found with a repulsive force, because the mass is shoved away and does not come back.

All the curves are interrelated because they are the orbits of masses moving in inverse-square force fields, and they are all related geometrically as conic sections.

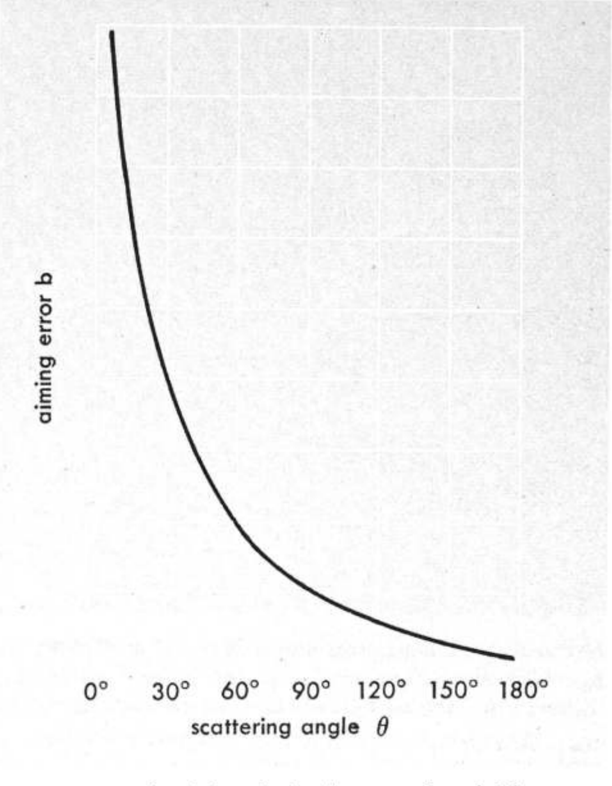


26–9 **Paths of a rolling ball on the hill model. The ball is sent with the same energy each time, but with a different aiming error. Note that the smaller the aiming error is, the larger is the scattering angle.**

26–3 ANGULAR DISTRIBUTION OF SCATTERING

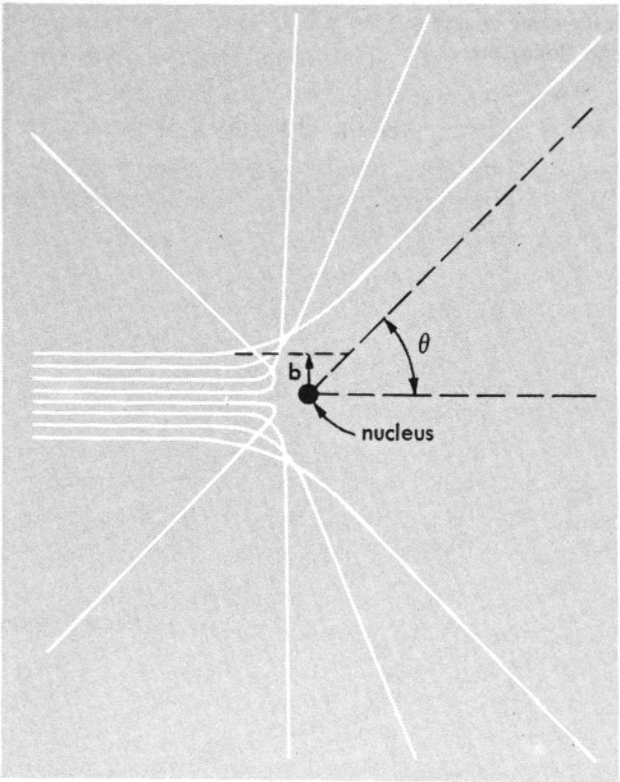
For particles of the same energy a definite aiming error leads to a particular angle of scattering under the influence of the Coulomb force from the nucleus. By computing the paths for a number of aiming errors, b (or by more elegant methods), we can plot a graph of the aiming error b against the resulting scattering angle θ. For the Coulomb force this relation is shown in Fig. 26 – 10.

Now if we could see the center of an atom and aim an alpha particle with complete precision so that we picked our aiming error b, we could test the relation between b and θ by a direct experiment. However, the alpha particles in the most accurately aimed beam will hit at random all over a region many times the area presented by an atom. Consequently, we must do a slightly less direct experiment.



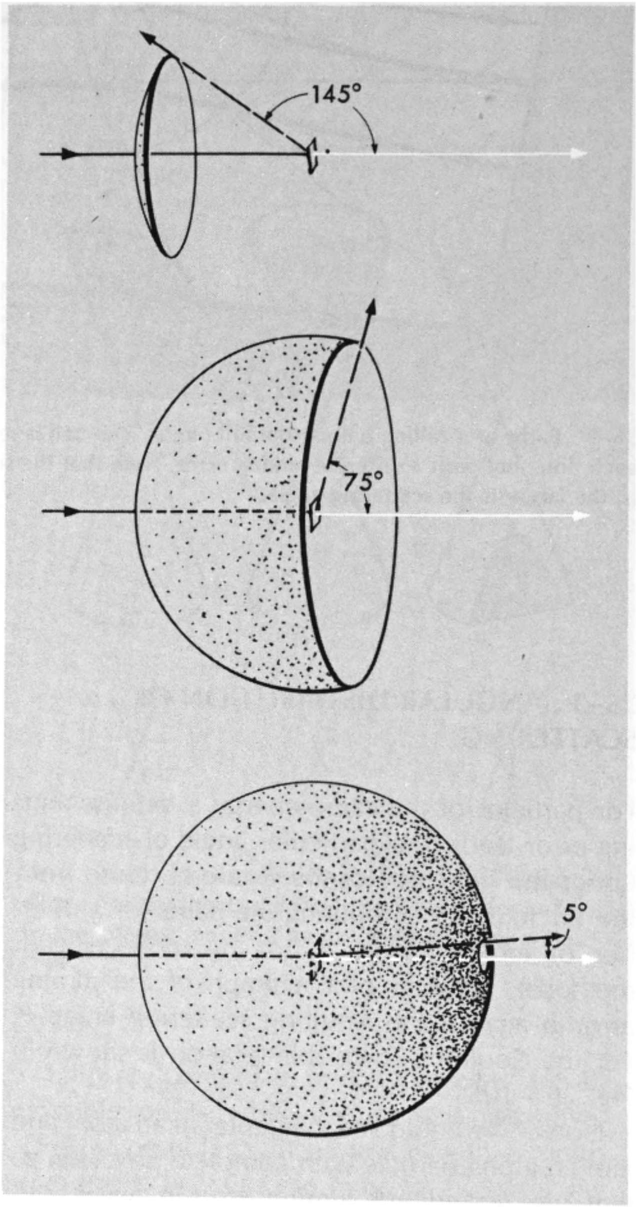
26–10 A graph of the relation between b and θ for a Coulomb force of repulsion.

Because smaller aiming errors lead to larger scattering angles (Fig. 26 – 11), we can take advantage of our uncontrollably poor aim by computing the number of alpha particles that will hit with less than a certain aiming error *b*. All these alpha particles (and no others) should then be scattered by more than the corresponding scattering angle θ. Since we can measure experimentally the number of alpha particles N_θ scattered by more than any particular angle θ, we can then find experimentally whether the scattering is produced by a Coulomb force. Another force might produce the same number N_θ scattered by more than one particular angle θ, but as we shall see, it could not produce the same relation between N_θ and θ over many angles. In detail, therefore, our experiment will be to measure N_θ as a function of θ (Fig. 26 – 12) and to compare the experimental results with those we predict from the Coulomb force.



26–11 **As the aiming error b decreases, the scattering angle θ increases. Therefore, all particles aimed within a distance b of the nucleus will be scattered through angles greater than θ. You see this same relation in Fig. 26–9 and 26–10.**

To compute the number of alpha particles that hit within a certain distance *b* of the center of an atom, we take advantage of the fact that the beam sprays uniformly all over the target area of an atom. Because the number of alpha particles aimed at a small area of an atom is equal to the number aimed at any other equal



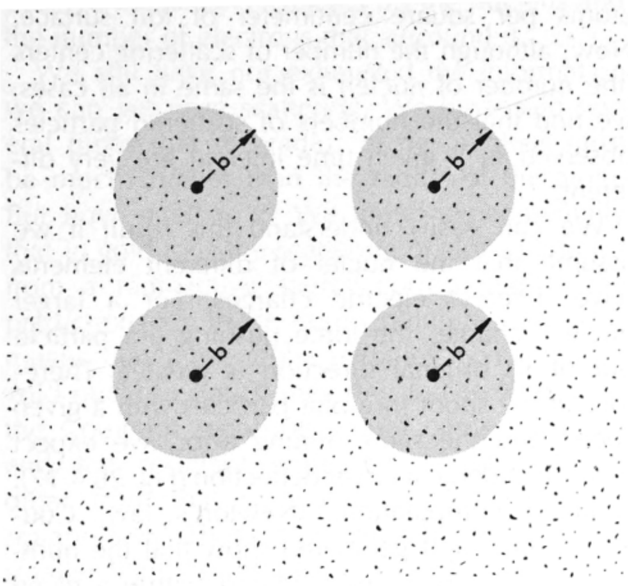
26–12 **Since the particles are scattered in all directions, they spread out over the surface of a sphere. When we count all those particles scattered by more than a given angle θ, we are counting the particles that fall on a cap, on a segment of the sphere. The caps for $\theta = 145°$, $\theta = 75°$, and $\theta = 5°$ are shown. N_θ is the number of particles scattered onto the appropriate cap.**

area in the same atom, the number of alpha particles that pass within a distance b of a given nucleus is proportional to the area of a circle of radius b (Fig. 26 – 13). This area is πb^2. Consequently, the number of alpha particles that are scattered through more than the angle θ is proportional to b^2.

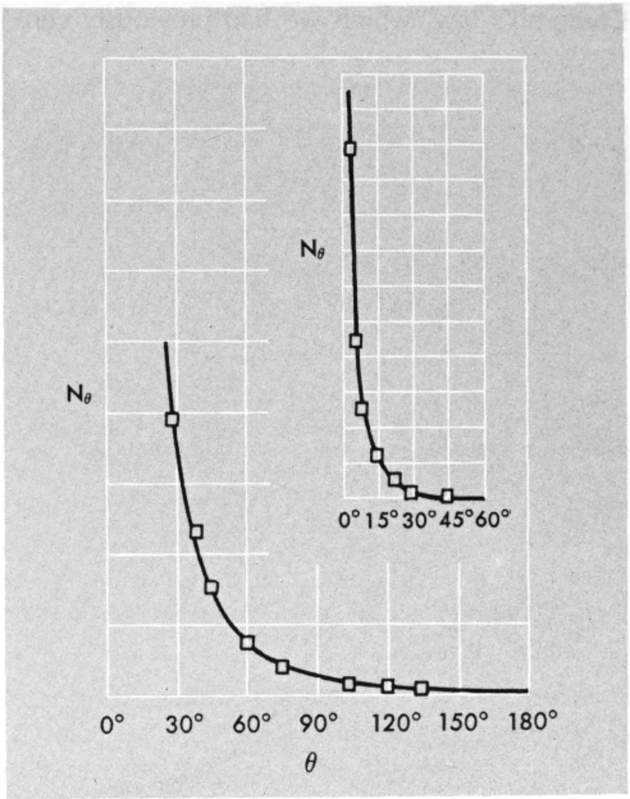
Now we already know how b is related to θ. By using the graph of b vs. θ we can construct a graph of b^2 vs. θ. Since the number of particles N_θ scattered through more than the angle θ is proportional to b^2, a plot of b^2 vs. θ and a plot of N_θ vs. θ are the same (except for a factor to adjust the vertical scale). Figure 26 – 14 shows this graph.

We are now in a position to make an experimental test of the Rutherford model. By counting the number of alpha particles scattered out of the beam into all various directions, we can find out experimentally whether or not the number of alpha particles N_θ—the number scattered through all angles larger than θ—looks like the graph. Detailed experiments on the angular distribution were performed by Geiger and Marsden (1913). The square boxes in Fig. 26 – 14 show some of their results. Within the accuracy of the measurements, the experimental points fall on the computed curve for the Coulomb force.

We can now conclude that the scattering of alpha particles is the result of a Coulomb force; but to be more certain of this conclusion, we shall see what kind of graph of N_θ versus θ would result from a different force. It is easy to calculate the paths and find the relation between aiming error and scattering angle for a force which is zero at large distances from the center of the atom and then rises abruptly to large values if an alpha particle gets within a certain distance of the center. For this reason we have chosen to construct the graph of N_θ versus θ for



26–13 Alpha particles are sprayed uniformly over the target area of many atoms. Therefore, the number of particles coming within a distance b of any nucleus is proportional to πb^2 – the area of a circle of radius b.

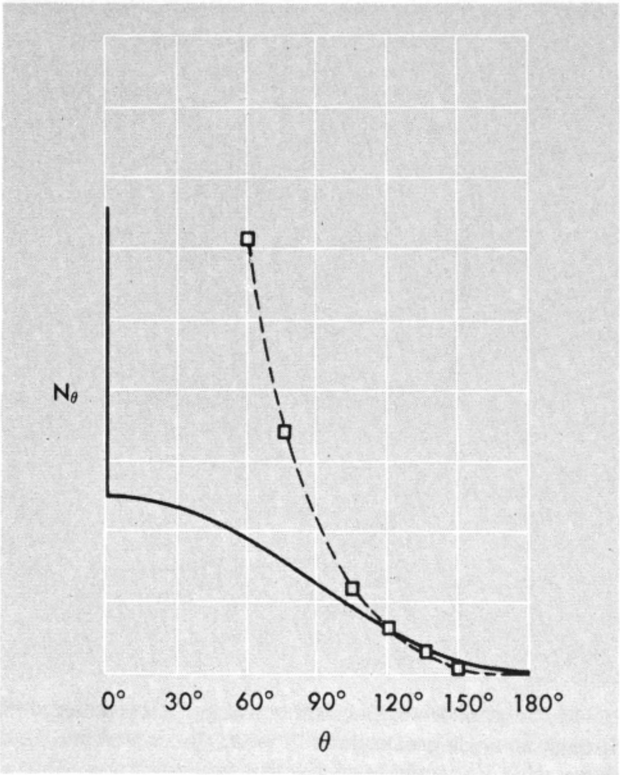


26–14 A graph of N_θ (the number of particles scattered through an angle greater than θ) vs. θ. The curved line shows what we would expect with a Coulomb force. The points in small squares represent the data collected by Geiger and Marsden in their scattering experiments. The insert shows the curve for small angles on a different scale.

such an abrupt force. This curve shown in Fig. 26 – 15 is different from the curve that results from a Coulomb force, so different that we clearly cannot get it to agree with the experimental results of Geiger and Marsden.

In general we may assume that any force we fancy is acting on the alpha particles; then we can compute the graph of N_θ versus θ for that force. For each different force, we get a different graph—and we can identify the force by the shape of the graph. The shape of the graph of the experimental results is the same as that for the Coulomb force. Consequently this is the force that acts on the alpha particles.

This is indeed a most remarkable result. Using alpha particles as probes, we have explored the electric field *inside* individual atoms and found that it obeys an inverse-square law. In the first place, this means that the positive charge of the atom is actually concentrated in a much smaller volume than the complete atom. In the second place, we find evidence that Coulomb's law, which we had previously verified experimentally only down to distances of the order of centimeters, holds in fact with high accuracy down to distances much smaller than atomic dimensions. As we shall see in the next section, those alpha particles with which we probe closest to the nucleus and for which we get the largest angles of deflection penetrate far inside the atoms. They only turn around when they are about 10^{-14} meter away from the nuclear charge, about one ten-thousandth of an atomic dimension. The known domain of the Coulomb force is thus extended by a factor of 10^4 toward the very small.



26–15 The solid line is the curve of N_θ vs. θ for a force that rises abruptly. The dashed line connects Geiger and Marsden's experimental points. The curves were adjusted so that they coincided at $\theta = 120°$. No adjustment will make them fit together.

26–4 MORE INFORMATION FROM SCATTERING

There is more information to be obtained from the scattering experiments. Suppose that we keep our source of alpha particles and our detector in fixed positions and we change to a leaf of a different metal (Fig. 26 – 16). Suppose that we use, one after the other, very thin sheets of different elements (gold, silver, platinum, etc.), choosing the thicknesses so that in each foil there are the same number of atoms in the way of the beam, that is, the same number of atoms per square centimeter of foil surface. Now, although the number of scattering centers (the number of nuclei) is the same in all cases, we find that the numbers of scattered particles observed in a given time interval are very different.

We can easily understand this result if we assume that the nuclei of different elements have different electric charges. For a larger nuclear charge, the force pushing the particle out of its original direction is greater. Therefore, if we shoot in alpha particles with a given speed and the same aiming error, we expect them to undergo a larger deflection (Fig. 26 – 17). Detailed computations (Newton's law, Coulomb's law, and geometry) show that the number of alpha particles scattered within a given range of angles is proportional to the square of the nuclear charge. Therefore, we can actually compare by our experiments the electric charges of the nuclei of different elements. We can go further still and compute the actual charge in elementary electric charges.

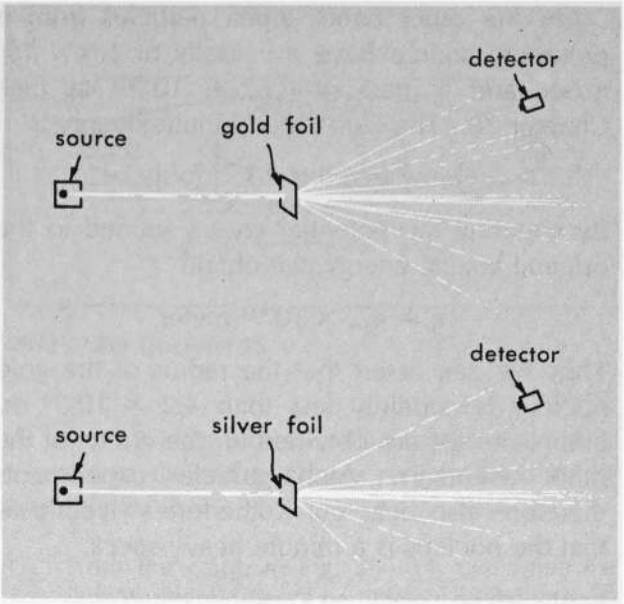
According to other evidence, the nucleus of copper has 29 positive elementary charges, the nucleus of silver 47, the nucleus of platinum 78. From alpha-particle scattering, Chadwick found almost these results. His experimental measurements yielded 29.3, 46.3, and 77.4 ± $1\frac{1}{2}$%. This means that the nuclear charges must be 29 or 30, 46 or 47, and between 76 and 79 elementary charges respectively—a masterpiece of prying information out of atoms with simple apparatus.

Also we know from other evidence that the nuclei of hydrogen and helium have one and two elementary charges respectively. (See Chapters 25 and 28.) Scattering experiments with hydrogen and helium gases (e.g., in a cloud chamber) confirm this result.

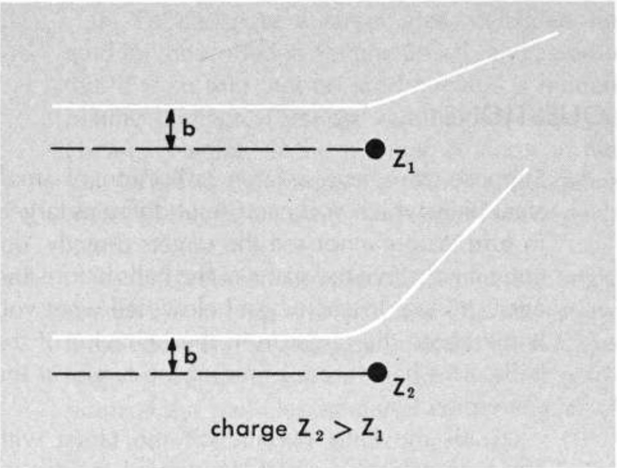
The number of elementary charges in the nucleus is called the *atomic number* of the element. Earlier in the development of chemistry this number was used for the serial number of the element in the chemical list. It is obviously a quantity of great physical significance. As we have mentioned, it can be measured by other methods as well as by alpha-particle scattering. They all agree and they give a new meaning to the serial numbers in the table of elements. (See Appendix 3) Because atoms are neutral, the atomic number tells both the number of positive elementary charges on the nucleus and the number of electrons that surround it.

We still have not exhausted the information we can get from scattering. We have said several times that the radius of the nucleus must be much smaller than the radius of the atom. But how small is it? We found experimentally that in gold foils the scattering of alpha particles from a radioactive source corresponds to the predictions we made from Coulomb's law. This means that these alpha particles never "hit" a gold nucleus. For, if they did, forces other than the Coulomb repulsion would come into play and the resultant scattering would certainly differ from that computed from the inverse-square law. Thus the gold nucleus must be smaller than the distance of closest approach of the alpha particles to the center of the nucleus.

Alpha particles of a given energy come closest to the nucleus when they are aimed directly at its center. (They move then on a straight line until their potential energy equals their initial kinetic energy, and then retrace their path backward. See Section 26 - 2.) Let r_0 be the distance of closest approach. Since the gold nucleus has 79 elementary charges and the alpha particle 2, the potential energy at the distance r_0 is



26–16 If we use foils of different metals, but each with the same number of atoms per cm² of surface, the observed scattering is different. Using the same beam for equal times, the number of alpha particles scattered into any given angle by the gold foil is 2.8 times the number scattered into the same angle by the silver foil.



26–17 Alpha particles of a given energy and aiming error are deflected more by nuclei of larger charge.

$$U = k\,\frac{(79\text{ elem. ch.})(2\text{ elem. ch.})}{r_0}$$

$$= (3.6 \times 10^{-26}\text{ joule-m})\,\frac{1}{r_0}$$

$$\text{since } k = 2.3 \times 10^{-28}\,\frac{\text{newton-m}^2}{(\text{elem. ch.})^2}$$

On the other hand, alpha particles from a polonium source have a velocity of 1.6×10^7 m/sec and a mass of 6.62×10^{-27} kg. (See Chapter 28.) Therefore, their kinetic energy is

$$\frac{1}{2}mv^2 = 8.5 \times 10^{-13} \text{ joule}$$

By equating the potential energy gained to the original kinetic energy, we obtain

$$r_0 = 4.2 \times 10^{-14} \text{ meter}$$

Thus we can assert that the radius of the gold nucleus is certainly less than 4.2×10^{-14} m. Similar results are obtained for the nuclei of the other elements. Alpha-particle experiments therefore also bear out Rutherford's hypothesis that the nucleus is a minute heavy speck.

QUESTIONS

1. Suppose you have a large collection of small steel balls which you can fire at different targets in turn. You cannot see the targets directly, but you can observe the paths of the balls before and after striking. In each case below, tell what you know about the target from the behavior of the balls. The balls are all fired at the target in the same direction.

   (a) All the balls bounce off the target with angle of reflection equal to angle of incidence.

   (b) All the balls bounce back, but with no definite relation between angle of reflection and angle of incidence.

   (c) Practically all the balls go through the target, but a few are lost in it.

   (d) All the balls are lost in the target.

   (e) The paths of all the balls cross after reflection at a point in front of the target.

2. To help visualize the thickness of the gold foil used in the alpha-particle deflection experiment described in Section 26 – 1, find the number of sheets of this foil which, when stacked together, are equal to one leaf of your textbook.

3. One square meter of silver foil has a mass of 5×10^{-3} kg.

   (a) What is its thickness? (The density of silver is 1×10^4 kg/m^3.)

   (b) The mass of a silver atom is 1.8×10^{-25} kg. What is the length of the side of a "box" representing a single atom?

   (c) What is the thickness of the foil expressed in atomic "boxes"?

4.* What is the evidence for our belief that alpha particles which are deflected by a large angle when passing through a thin gold foil do so in one collision?

5.* What are the main features an atomic model must have in order to account for the way in which alpha particles are deflected when they pass through gold foil?

6.* What is a scattering angle?

7. Suppose that a ball is rolled many times at the hill model of Fig. 26 – 6. Each time it starts along the same path, but with a different kinetic energy. How do you expect its scattering angle θ to be related to its energy?

8. In a scattering experiment using gold foil, 4 alpha particles in 10,000 are scattered through an angle of more than 5°.

   (a) How many particles would be scattered through more than 5° if we doubled the thickness of the foil?

   (b) If we made the foil a thousand times thicker?

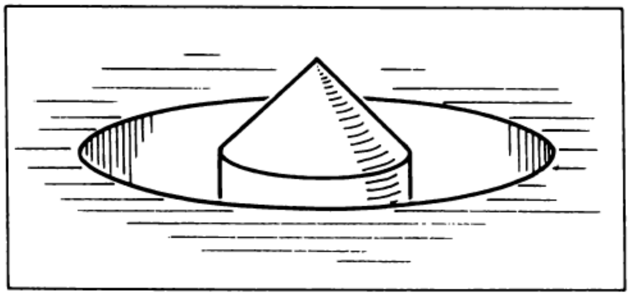
9. A beam of alpha particles from polonium is shot perpendicularly into the gold foil described in Section 26 – 1. We observe that some of the particles are scattered through an angle of about 3° or 5×10^{-2} radian. The alpha particles have a kinetic energy of 8.5×10^{-13} joule, which corresponds to a speed of 1.6×10^7 m/sec.

   (a) What velocity $\Delta\vec{\mathbf{v}}$ must be added at right angles to their original velocity to produce this change in direction?

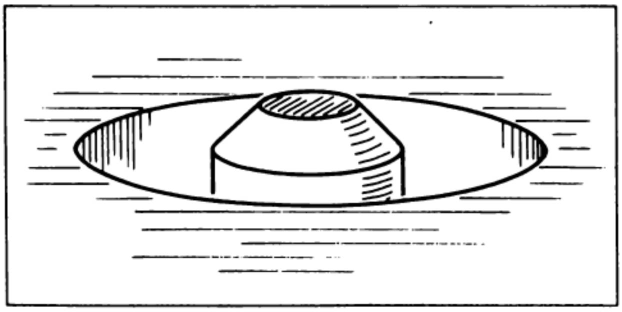
   (b) How long does it take an alpha particle to pass through a single gold atom in the foil?

   (c) What is the average acceleration of one of these alpha particles if it gets its whole deflection in passing through a single gold atom?

10.* Why is it not possible to find directly the relation between θ and b from the Rutherford scattering experiment with alpha particles?



26–18 For Question 14.



26–19 For Question 15.

11.* How is N_θ, the number of particles which have scattered through all angles larger than θ, related to the aiming error b?

12.* What do the results of scattering experiments tell us about the electric field inside the gold atom?

13.* If the alpha particles approaching a nucleus with a certain aiming error b were scattered by 90°, through what range of angles were the particles scattered which approached the nucleus with a smaller aiming error?

14. Imagine a device which can drop 1000 steel balls through a circular hole 5 cm in radius and spread them evenly over the area of the circle (Fig. 26 – 18). In the center of the hole we fix a steel cone whose sides make an angle of 45° with its base, and whose base has a radius of 2 cm. In a group of 1000 balls, how many are scattered sideways before they reach the hole?

15. The point of the cone in Question 14 is sawed off to leave a smooth plane parallel to the base and 1 cm in radius (Fig. 26 – 19).

(a) In each group of 1000 balls, how many are turned back and how many are scattered sideways by it?

(b) Draw a graph of the scattering angle θ versus the aiming error b, measured from the center of the hole.

(c) Draw a graph of N_θ (the number of balls scattered through an angle greater than θ) vs. θ.

16. (a) To construct the curve of b vs. θ for an abrupt force, draw a fairly large circle to represent a plane cut through a rigid sphere and measure the scattering angles for various aiming errors. (Assume that the angle of reflection of a ball hitting the surface equals the angle of incidence. Why?)

(b) From your curve of b vs. θ construct a graph of N_θ vs. θ and compare it with Fig. 26 – 15.

(c) Explain why this relationship is independent of the energy of the particles fired at the sphere.

17. From the graph in Fig. 26 – 14, determine the ratio of the number of particles scattered through an angle greater than 15° to those scattered through an angle greater than 22.5°.

18. To make a practical calculation of the number of alpha particles scattered by a metal foil, we need to know the target area of each atom in the foil, and the way in which the angle of scattering depends on the aiming error.

(a) To calculate a target area, consider the gold foil described in Section 26 – 1, and assume that it is so thin that no gold nucleus is hidden behind another. Then the number of scattering centers is equal to the number of atoms in the foil. What is the number of atoms per square meter of this foil? What is the target area of each atom?

(b) The relation between the scattering angle θ and the aiming error b for a Coulomb's-law force between the alpha particle and the nucleus is given by the graph in Fig. 26 – 10. For polonium alpha particles scattered by the nuclei of gold atoms, the scale for the b axis is 2×10^{-14} meters for each unit on the graph. If a beam of polonium alpha particles falls on the gold foil of part (a), what fraction of the particles is scattered through more than 60°?

19.* How many times as close can an alpha particle of given energy get to the center of the nucleus of an aluminum atom as it can to the center of a gold nucleus?

20.* How close can polonium alpha particles get to the center of an aluminum nucleus?

21.* Which conservation law is used in calculating the distance of closest approach?

22. Three experiments were done with the same alpha particle beam, the same counters at the same positions, and different foil targets. Five counters were arranged with respect to each target and the beam to count particles scattered by the angles given in the table. The time interval of counting was the same for all three experiments. The foils were designed so that the beam passed through the same number of atoms per square centimeter regardless of which target was used. The results presented in the table below give the numbers of particles scattered into the same range of scattering angles. For example, the count at 30° includes all particles scattered in any direction in the range between 29° and 31°; the count at 60°, all those scattered in any direction in the range between 59° and 61°, etc.

SCATTERING ANGLE	NUMBER COUNTED		
	Target 1	Target 2	Target 3
30°	2790	35,920	102,810
60°	346	4451	12,760
90°	100	1288	3685
120°	39	496	1423
150°	14	187	532

(a) Plot the number scattered versus the scattering angle for each target. (Plot all three graphs on the same axes.)

(b) Target 1 was aluminum. What were the other targets?

23. (a) Explain why the gravitational force between an alpha particle and a gold nucleus has little effect on the scattering of alpha particles by gold.

(b) What is the ratio of the electric force on an alpha particle to the gravitational force when the separation of an alpha particle and a gold nucleus is 10^{-14} m? When it is 10^{-13} m?

24. An alpha particle of kinetic energy 1.0×10^{-12} joule is scattered by a gold nucleus.

(a) What is the minimum possible distance of closest approach?

(b) What must be the aiming error if the alpha particle is to get this close?

(c) How close could a proton of the same energy approach? A neutron?

25. In one scattering experiment a proton beam strikes a gold target. In a second experiment a beam of deuterons (atomic mass = 2, atomic number = 1) is used. The kinetic energy of the particles is the same in both beams. How will the minimum distance of closest approach to gold nuclei of the two kinds of beams compare?

26. (a) What fraction of its energy is lost by an alpha particle in a head-on collision with an electron? (The mass of an alpha particle is about 7200 times that of an electron.)

(b) Approximately how many such collisions must occur before the energy of the alpha particle is reduced by 1 percent?

27. An alpha particle with speed 1.00×10^7 m/sec is scattered by a gold nucleus at rest through an angle of 60°. Assume that so little kinetic energy is transferred to the gold nucleus that the speed of the alpha particle after the scattering is the same as it was before.

(a) What is the momentum transferred to the gold atom?

(b) What is the velocity given to the gold nucleus?

(c) What is the kinetic energy given to the gold nucleus?

(d) Was your initial assumption a good one? Explain.

28. (a) If the distance of closest approach of the alpha particle in Question 27 was 1.65×10^{-13} m, what was its speed at this point?

(b) At what speed was it moving toward the nucleus at that point?

29. In the hydrogen atom, the electron is on the average 5×10^{-11} m from the positively charged proton.

(a) Estimate the average force between the electron and proton.

(b) If the atom were placed between the plates used in the Millikan experiment of Section 24 – 7, how many of the 90-volt batteries would we need to put in series with the plates to tear the atom apart?

27

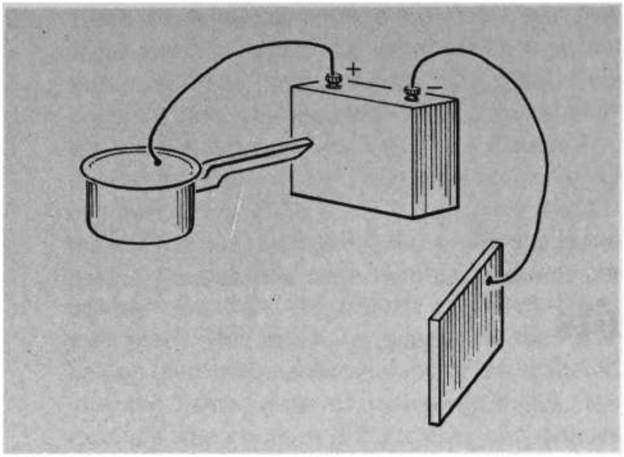
Electric Circuits

27–1 CONDUCTORS, BATTERIES, AND POTENTIAL DIFFERENCE

In this section we shall make several applications of the idea of electric field and especially of electric potential difference. In particular we shall consider the fields and potentials of electrical conductors both when they are isolated and when they are connected to batteries.

In charging a conductor, electric charges move; but as we learned in Secton 24 – 3, after charges have rearranged themselves (in response to electric forces exerted by the charged bodies in the neighborhood), all the electric field lines meet the surface of the conductor at right angles. Otherwise charge would move along the surface until the component of electric field in that direction was eliminated. Also there is no electric field inside the conductor, for if there were, charges would be moved by it until the field inside was canceled. What, therefore, is the potential difference between two points either on the surface or inside of a charged conductor? Since there is no electric field either along the surface or through the inside of the body, the force on an elementary charge is zero all along any path passing from one point to the other through the conductor. An elementary charge can therefore be moved without any expenditure of energy. The conductor is all at the same electric potential; there is no potential difference between any of its points. A body which is at a single potential (or a surface all at the same potential) is called an equipotential. In the vocabulary of potential difference we can, therefore, say that any conductor in which charges do not move is an equipotential body.

We have considered what happens when a battery drives charges around an electrical circuit. Now let us consider what happens when the circuit is open. For example, suppose we connect conductors to each battery terminal as shown in Fig. 27 – 1. Then the battery will move charges, concentrating negative charge on the conductor connected to the negative terminal and positive charge on the conductor connected to the positive terminal. The charges will move as long as the energy required to carry one positive elementary charge from the negative to the

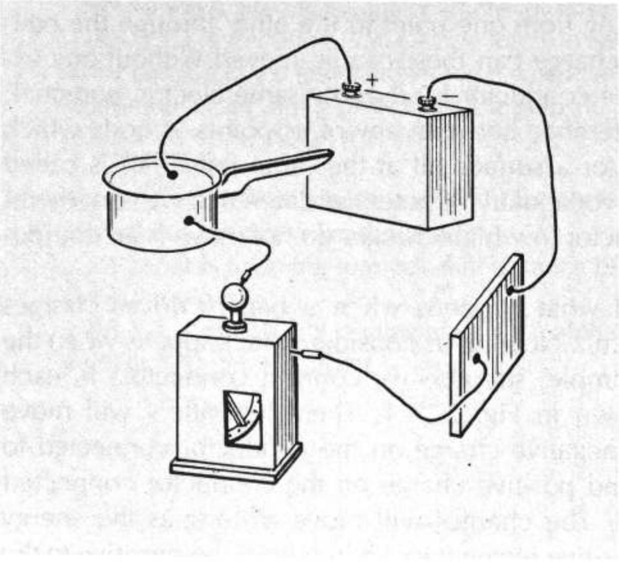


27–1 A battery connected to two conductors.

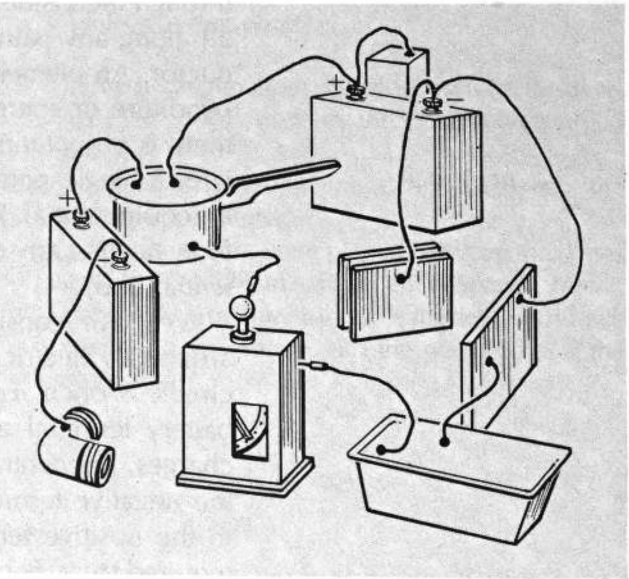
positive conductor is less than the energy supplied to each elementary charge by the battery. As the charge concentrations are increased, however, any positive charge moved between them is attracted to the negative conductor and repelled from the positive conductor with increasingly greater force. Consequently, as the battery charges the conductors, the energy required to move an elementary positive charge rises. And if the charging process went on indefinitely, the energy required to move a single elementary charge between the conductors would eventually exceed the energy the battery can supply to each elementary charge. Hence, the charges stop moving, and the charge distribution stops changing, when just enough charge has been transferred so that the energy *required* to move an elementary charge is equal to one elementary charge times the EMF of the battery.

The energy per elementary charge required to move a positive elementary charge from the negative conductor to the positive conductor is exactly what we mean by the electric potential difference (or potential drop) from the positive to the negative conductor. When the conductors are charged up, the potential difference between them is the same no matter what points we select on the conductors, for we already know that each conductor is then an equipotential body. Also, when the conductors are connected to the battery, they are always charged to the same potential difference, equal to the EMF of the battery, no matter how large the conductors are or where they are located. We can measure this potential difference with an instrument called an *electrometer* [Fig. 27 – 2 (a)], which will be explained in the next section. We can hang as many conductors from the terminals of the battery as we wish without influencing this result. Even connecting another battery as in Fig. 27 – 2 (b) does not change the potential difference maintained by the first one. In summary, then, when no current is flowing through a battery, we can view the battery as a machine for maintaining a definite difference in potential, the potential difference given by its EMF.

A conductor in which charge has stopped flowing is an equipotential body. But what of a conductor such as a wire connected to the terminals of a battery? Charge flows steadily



27–2(a) The potential difference between any two conductors connected to a battery, as measured by an electrometer, depends only on the EMF of the battery.



27–2(b) The electrometer gives the same reading as in (a), even though one of the conductors is connected to another battery and more conductors have been added.

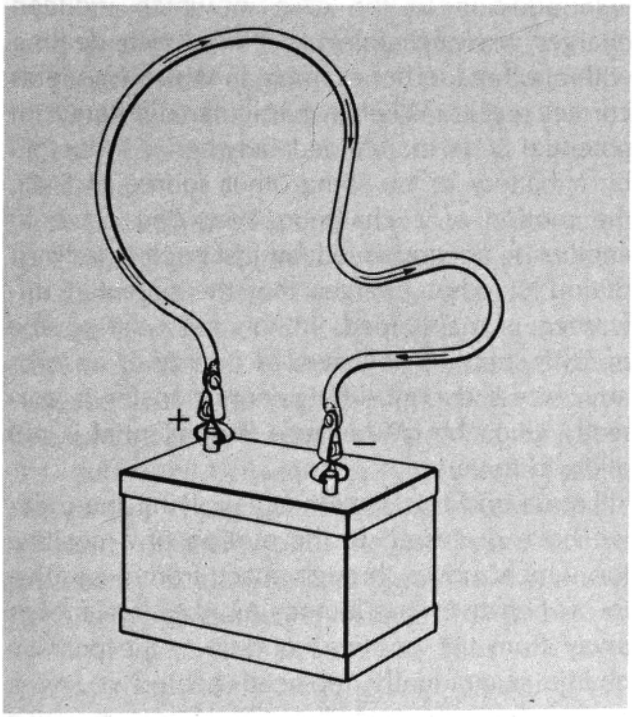
through it, driven by an electric field. To find the complete pattern of the field is a more complicated job than we wish to do here, but we can tell a good bit about the field nevertheless. First, the field lines inside the conductor cannot run into its surface. If they did, charge would flow onto the surface until a canceling field was built up. Indeed, charge concentrations on the surface are built up as soon as the circuit is connected. From then on the field inside the wire points from the positive battery terminal to the negative terminal. No matter how many bends there are in the conductor, inside it the field lines will follow around each curve (Fig. 27 – 3). This field inside is, therefore, relatively simple compared with the field outside the conductor. The field lines outside can run in any direction from the charges on the surface. Unlike the field inside, the field outside depends on the location of the battery, the shape of the conductor, and the position of the other charged bodies.

The direction of the electric field inside a conducting wire points parallel to the sides of the wire. The magnitude of $\vec{E}$ must be constant throughout a uniform wire in order to drive the same current through all sections, as we shall see in Section 27 – 4. Hence, the work done per elementary charge as a positive charge moves from one end to the other of a wire of length l equals the magnitude of the force per elementary charge times the distance moved, or El. Since the battery supplies the energy which is given to the charge, the work done per elementary charge must equal the energy supplied per elementary charge by the battery, or $\mathcal{E}$. That is,

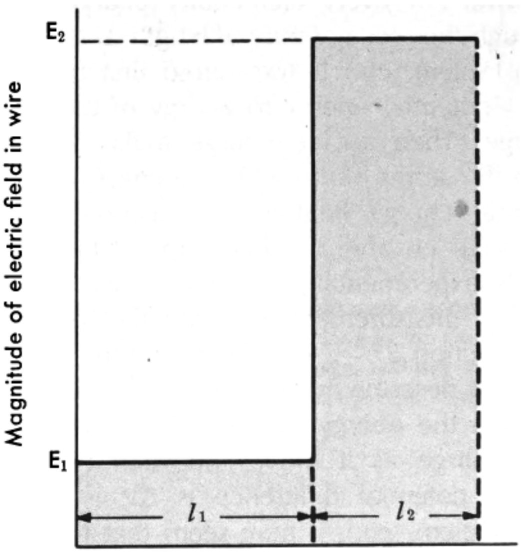
$$\mathcal{E} = El$$

For a given battery connected to the ends of the wire, $\mathcal{E}$ is constant, so the magnitude of the electric field inside the wire depends on the length of the wire.

If the wire is not uniform, the magnitude of $\vec{E}$ varies from point to point. But if we plot E against the distance along the wire from the positive battery terminal, the area under the curve between the end points of the wire is equal to the energy supplied per elementary charge to a positive charge moving the entire distance; that is, it is equal to the potential difference between the terminals. For instance (Fig. 27 – 4), if the wire is a good conductor for the length l_1 and a poor conductor for the length l_2, charge will accumulate at the junction between the two conductors until the field E_2 in l_2 is sufficiently stronger and the field E_1 in l_1 is sufficiently weaker so that the same current flows in each. The sum, $E_1l_1 + E_2l_2$, is then the potential difference between the positive and negative battery terminals. This is the total



27–3 When a conducting wire is connected to a battery, there is an electric field inside the wire that points from the positive to the negative battery terminal.



27–4 A graph of the magnitude E of the electric field in a wire vs. the distance l along the wire. Here the wire is a good conductor for one part of its length and a poor conductor for the other. The area El under part of the curve gives the energy transferred per elementary charge in this length. The total area gives the potential difference between the two ends of the wire.

energy per elementary charge supplied to charges passing through the wire between the terminals and it is determined by $\mathcal{E}$, the EMF of the battery. If it were less than $\mathcal{E}$, the battery would quickly increase it by forcing more positive charge onto the positive terminal and more negative charge onto the negative terminal; if it were more than $\mathcal{E}$, the battery would be prevented from separating charge until the terminals discharged and the energy per elementary charge required to force charges onto them again became $\mathcal{E}$.

We have just seen that the EMF of the battery equals the total potential difference, $E_1l_1 + E_2l_2$, between the positive and negative battery terminal. That is,

$$\mathcal{E} = E_1l_1 + E_2l_2$$

This potential difference appears on the right-hand side of the equation in two parts: E_1l_1 is the potential difference across the good conductor of length l_1, and E_2l_2 is the potential difference across the poor conductor of length l_2. In fact as we go along the conductor from the positive battery terminal to the negative terminal, the potential difference measured from the positive terminal to any point on the wires changes continuously. From one end to the other of any little length of wire Δl the potential changes by the amount $\Delta V = E\Delta l$ where E is the magnitude of the electric field in this piece of wire. For every elementary charge passing through the piece of wire of length Δl the energy (ΔV)(1 elem. ch.) is transferred first from electrical potential energy to energy of the moving charge. Then as the charge makes collisions with the atoms of the wire, the energy goes into thermal energy heating that section of wire. We shall be able to check up on these statements experimentally as soon as we have discussed measurement of potential difference. (See Section 27 – 2.) We then find that our conclusions describe the experimental facts.

Since the energy transferred to one elementary charge as it moves between two points whose potential difference is ΔV is equal to (ΔV)(1 elem. ch.), it may seem that the energy transferred to any charge q is $(\Delta V)q$. Often this conclusion is correct. But it depends on the assumption that the neighboring charges which produce the electric field stay in the same positions and thus exert the same forces that they do when we move only a single elementary charge. Remember that in the last chapter we specified that if the charges producing the field were nailed down, the energy would indeed be $(\Delta V)q$. However, if these charges are on a conducting body, it is impossible to nail them down; they will move under the influence of the Coulomb forces exerted on them by q. Even a single elementary charge will induce a redistribution of charges when it moves from one place to another in the neighborhood of a conducting body. In fact, when a charged body is brought near an *uncharged* conductor, a redistribution of charges takes place and the charged body experiences an attractive force [Fig. 27 – 5 (a)]. On the other hand, if the conductor has a great enough excess of charges, the change in the electric field due to induced charges will be negligible [Fig. 27 – 5 (b)]. In that case, we can use the potential due to the charges on the conductor, neglecting the redistribution of charges, to give the change in potential energy as $(\Delta V)q$.

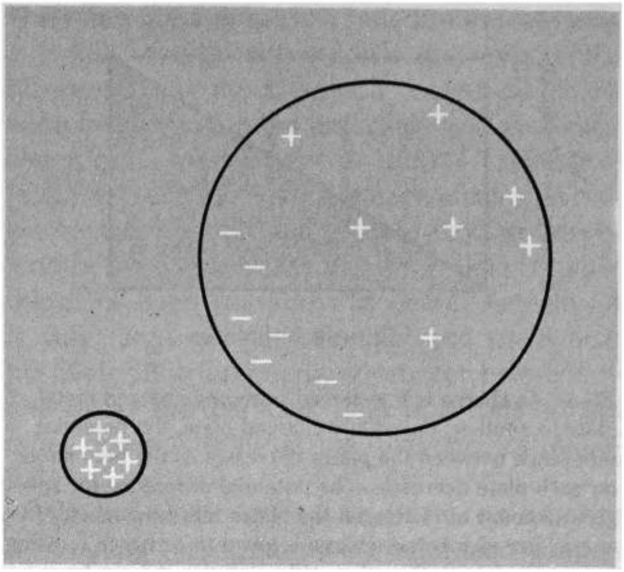
The potential V at a point is really the potential difference between that point and infinity, since we have chosen the potential at infinity to be zero. From now on we shall use the symbol V alone, rather than ΔV, to represent the potential difference between any two points, whether or not one of them is at infinity.

In addition to the case in which induced charges are negligible, we have been dealing with another kind of example in which qV yields correct results. When a particular difference of potential V is maintained between conductors by a battery or by some other source of EMF, the motion of a charge q from one place to another is accompanied by just such a redistribution of other charges that the potential difference is maintained. In this case the source of EMF supplies any needed energy or absorbs any excess energy. The energy transfer is correctly given by qV because the potential is not allowed to vary.

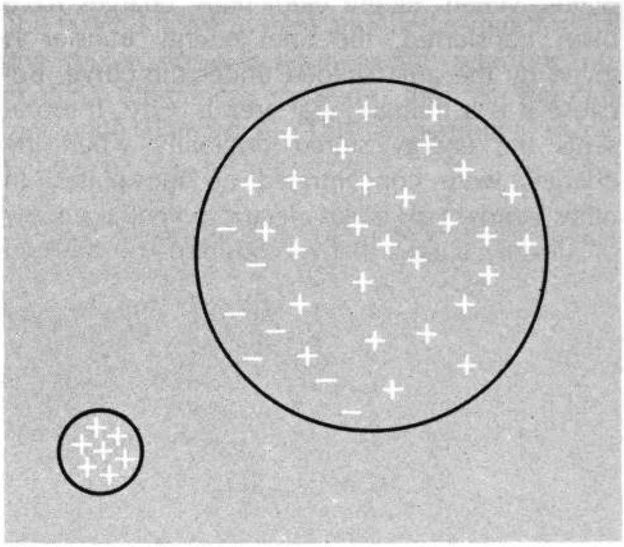
Let us end this section by applying the ideas we have discussed to the motion of a positive ion which moves through space from a positive to a negative conductor. As the ion moves away from the positive conductor, the positive charge is gradually replaced behind it. As it approaches the negative conductor, negative canceling charge gradually concentrates directly in front of it. If the two conductors are connected through a battery, positive charge is gradually transported through the battery from the negative plate to the positive plate. Thus at the end of the process the canceling negative charge has been supplied to the negative conductor, and a replacement positive charge has been given to the positive conductor. Also, as the charge is transported through the battery, the battery supplies an amount of energy equal to that which goes into the kinetic energy of the positive charge moving between the conductors. When this whole process is finished, the situation on the conductors is the same as it was at the beginning. Only the chemical energy of the battery and the thermal energy of the conductors change. The next charge that goes across between the conductors will get exactly the same energy as the first one, and if charges whose total charge is q go across, they will receive the energy qV.

On the other hand, if there is no battery, the positively charged conductor permanently loses the charge of the ion, and the negative conductor has one of its negative charges permanently canceled. The potential difference between the positive and the negative conductor drops because of the reduction in charge. The next charge to cross between the conductors will, therefore, be given somewhat less energy, and each successive charge crossing will get a progressively smaller amount. Consequently, the total charge q crossing receives less than the energy qV, where V is the original potential difference.

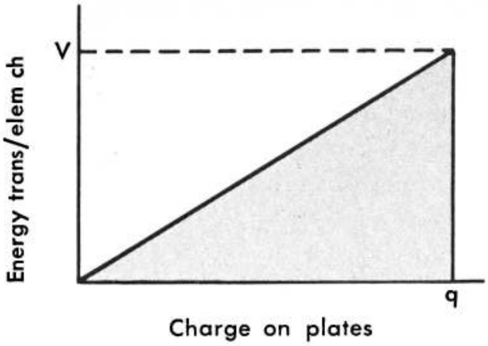
As an extreme example, we consider two oppositely charged metal plates. As elementary charges pass between them, the plates discharge; the potential V also goes to zero, staying proportional to the charges on the plates. Consequently, although the first charge receives the energy (1 elem. ch.) V, where V is the potential difference between the fully charged plates, successive charges receive smaller and smaller amounts of energy until the last charge receives no energy at all. The average amount of energy received per elementary charge is just half the original potential difference. Thus the total energy received by a total charge q passing from one plate to the other is $\frac{1}{2}qV$. You can see that



27–5(a) If we bring a charged body near an uncharged conductor, it induces a distribution of the charge. These induced charges exert a force on the charged body, showing that there is an electric field from the induced charges.

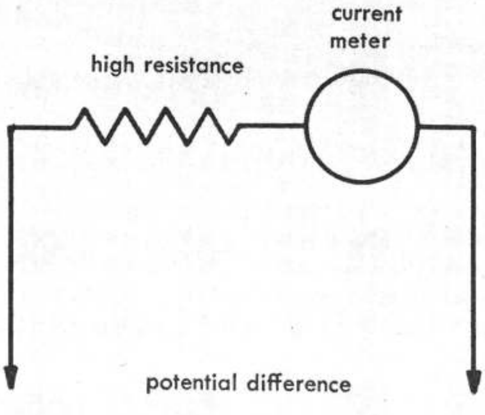


27–5(b) A charged body is brought near a conductor with a large excess of charges. As long as the two bodies are not too close, the induced charges will not noticeably change the original electric field of the conductor.



27–6 As charge is transferred from one charged metal plate to another, oppositely charged plate, the potential difference between the plates decreases as the net charge on each plate decreases. The potential difference for any given amount of charge on the plates tells how much energy per elementary charge is given to a charge crossing from one plate to the other. When all the charges have been transferred, the total energy they have received is given by the shaded area under the curve.

this must be true by looking at Fig. 27 – 6 where we have plotted the energy transferred per elementary charge vs. the charge then on the plates. When all the elementary charges have been transferred, the total energy transfer is given by the area shaded under the curve. Because it is a triangle, that area is $\frac{1}{2}qV$. It represents the energy stored originally when the charges were concentrated on the plates. In other words $\frac{1}{2}qV$ is the electric potential energy of all the charges in their positions on the two plates. Clearly this total electrical potential energy is to be distinguished from the electric potential difference V and from the energy qV that would be given to a total charge q passing between the plates if the potential were maintained by a battery.



27–7 A voltmeter to measure potential differences is made by connecting a high resistance and a current meter in series. Often the high resistance is included inside the case containing the meter.

27–2 MEASUREMENT OF POTENTIAL DIFFERENCE

We can use a shielded electrometer to measure potential differences (Fig. 27 – 2). An electrometer gives a particular reading when the forces on the stem have a particular value; and these forces are determined by the charge on the stem and the induced charge on the inside of the shield. For a particular charge on the stem there is a definite electric field in between the stem and the inside of the shield, and the electric potential difference has a particular value. The reading of the electrometer corresponds to this value of electrical potential difference. Conversely, whenever the electrical potential difference between stem and shield has the same value, the fields and charges inside the electrometer must be the same, and the electrometer will give the same reading.

We can measure potential difference with an electrometer. Often, however, it is more convenient to measure potential difference by using a current meter and a high-resistance wire in series (Fig. 27 – 7). When we apply various differences of potential V across the ends of this circuit, the electric fields driving charges through the circuit are proportional to the differences in potential. For example, if the difference of potential is maintained by concentrations of charge on conductors at the two ends of the circuit, the fields and the potential difference are proportional to those charge concentrations. Consequently we see that a higher potential means larger forces on the charges which are free to move through the high-resistance wire and the current meter. These larger forces lead to bigger currents through the circuit; and the current that we read is a measure of the potential difference. With ordinary resistance wires, as we shall see in Section 27 – 4, the current is actually proportional to the potential difference, but in any case we can calibrate the current meter with its high-resistance wire against known potential differences maintained by batteries. Once we have calibrated the current meter with its high resistance, it becomes a potential-difference meter. Such meters are usually known as voltmeters.

Voltmeters are convenient to use as long as the resistance in them is so high that the current through the voltmeter is small compared to any other currents we would like to maintain in the circuit, or as long as we can make a potential-difference measurement in a time so short that the current does not appreciably change any charge concentrations that we wish to maintain. When these conditions are not met, you must use more complicated circuits, for instance the circuits known as potentiometers which you can explore in the laboratory.

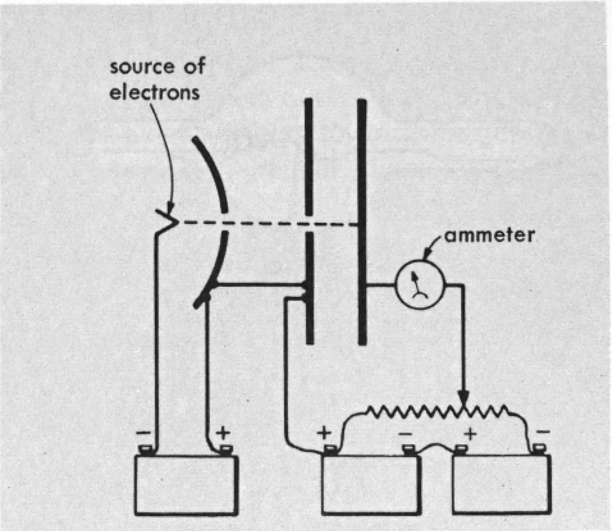
27–3 A FURTHER CHECK ON POTENTIAL DIFFERENCE AND ENERGY

A battery maintains a constant potential difference. In the absence of currents any point on a good conductor is at the same electrical potential. Consequently by maintaining a fixed potential difference between any two conductors with a battery, we can accelerate charged bodies to predetermined kinetic energies in the electric field between the two conductors. The conductors need not be large parallel plates.

We can allow some current in the conductors without losing our knowledge of the potential difference between them. Because charges move freely in the conductor practically no energy is required to move a charge from one point to another, and the conductor may be approximately at a single potential even though the charges are forced to move. (We shall make these considerations more quantitative in Section 27 – 5.)

We can apply these ideas to get a check on our knowledge of potential difference and energy. For example, we can compute the energy of the electrons emitted by the electron gun described in Chapter 23. This gun consists of a hot filament which boils out electrons with negligible speeds, and a plate which is maintained at a constant difference of potential with respect to the filament. The electrons are accelerated between the filament and the plate. Then they pass through a hole in the plate with a kinetic energy proportional to the potential difference. A thousand-volt potential difference between filament and plate will give each electron a kinetic energy of $1000 \times 1.6 \times 10^{-19}$ joule, because the charge of each electron is one elementary charge. The electric field in the region between the filament and the plate is something we need not know in detail. Actually, it is very large near the filament and small near the plate. But the energy given per elementary charge as the particles pass from one to the other depends only on the difference of potential.

We can check up experimentally on this fact by running the electrons from the electron gun into a uniform retarding electric field between parallel plates where we know the retarding force (Fig. 27 – 8). The electrons will go across the space between the parallel plates as long as the retarding potential difference between the parallel plates is smaller than the accelerating potential difference between the filament and plate of the electron gun. They stop reaching the far end of the system and the current through the meter stops as soon as the retarding potential difference exceeds the accelerating potential difference. This experiment shows



27–8 An experiment to show that the energy gained by charged particles depends only on the potential difference through which they move and not on the shape of the charged plates. The current through the ammeter stops whenever the potential difference measured across the plates becomes greater than the potential difference across the gun.

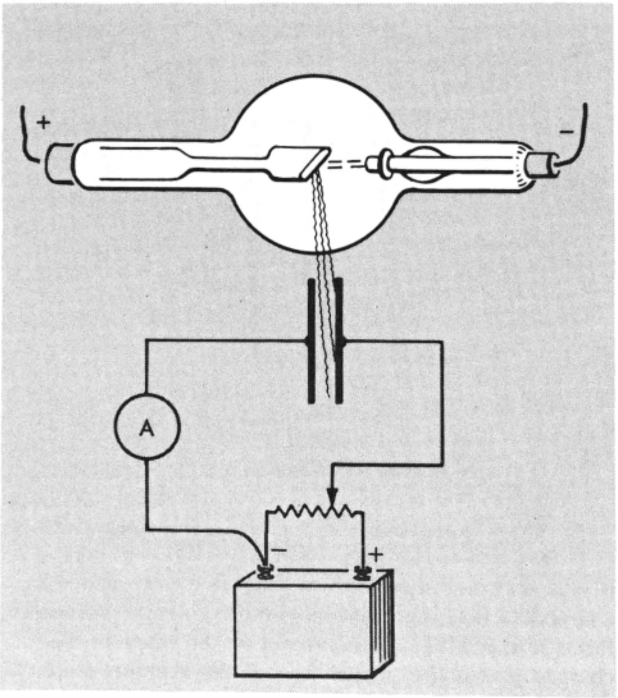
that it is the potential difference imposed by the batteries which determines the energy; the shape of the conductors is irrelevant.

27–4 CURRENT VERSUS POTENTIAL DIFFERENCE

How do the currents in ionized gases, the currents of electrons boiled out of metals, and the currents in metals depend on potential differences maintained across these circuit elements? This is the question we shall now examine.

A. Ionized Gases. We shall begin with the case of ionized gases. The experimental arrangement is shown in Fig. 27 – 9. Two parallel metal plates are connected to a variable voltage supply, and a sensitive current meter is inserted in one of the connecting wires. The plates are at a distance of several centimeters, and the space between is filled with the gas under investigation. When ionizing radiation (X rays or rays from a radioactive source) passes through the gas, the meter shows a current.

Figure 27 – 10 (a) shows how the current depends on the potential difference between the metal plates in one typical experiment. The gas is argon at atmospheric pressure, and the radiation comes from a sample of radium. As the potential difference between the plates is increased, the current increases rapidly at first, then more slowly, and eventually reaches a certain limiting value called the saturation current.



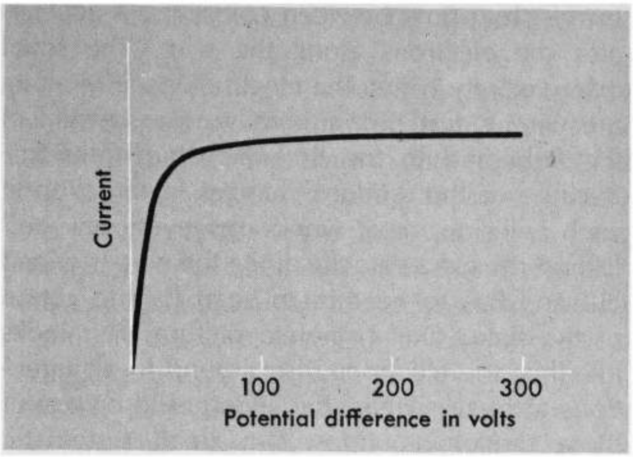
27–9 **Apparatus for investigating electric currents in ionized gases.**

The saturation current is reached when all the ions formed are swept out of the gas by the electric field and collect on the appropriate plates. For singly charged ions, such as those produced in argon, the ionization current (in elementary charges per second) equals the number of positive ions per second reaching the negative electrode (and this is the same as the number of negative ions—in this case electrons—reaching the positive electrode). Therefore the saturation current measures the total number of pairs of ions produced every second in the gas between the plates.

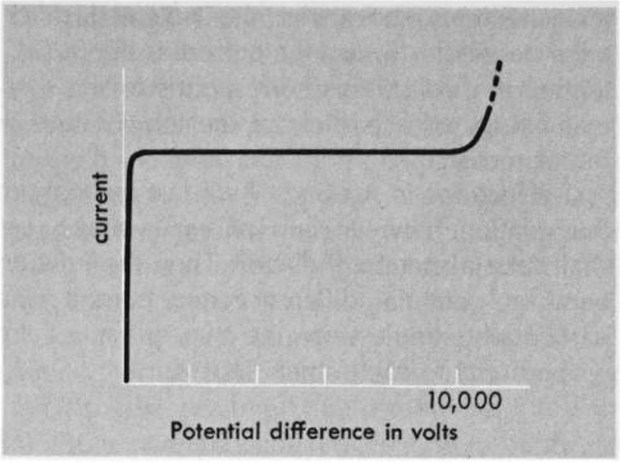
Why is the current less when there is a small difference of potential between the plates? The field moving the ions toward the plates is then weaker, and the ions move toward the plates more slowly. As they gradually get pushed toward the plates, they move in all directions because of their random thermal motions, and sometimes a positive and a negative ion will collide. Occasionally they will even recombine to form a neutral atom. When the systematic motion of the ions toward the plates is very slow, many positive and negative ions recombine instead of reaching the plates.

If we increase the potential difference between the plates far beyond the value necessary to establish saturation, we find that the current eventually begins to increase again. Then with a little more voltage it rises to very large values [Fig. 27 – 10 (b)]. At these potential differences, the gas between the plates glows and there are often crackling noises. Just what we observe at this stage varies greatly with the pressure, the nature of the gas, and the distance between the electrodes. At low pressures the current and the glow are more or less constant. At atmospheric pressure, on the other hand, we have a series of sparks.

These *gas discharges* are quite complex; but we can understand their fundamental cause. The electric field accelerates the positive and negative ions of an ionized gas in opposite direc-



27–10(a) Current vs. potential difference curve for argon gas at low voltages.



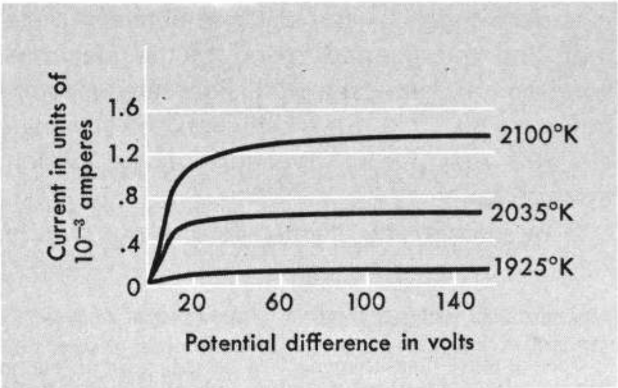
27–10(b) Current vs. potential difference curve for argon up to a high voltage.

tions. Each ion picks up energy until it collides with a gas molecule. On the average, in these collisions, the ions then lose the extra kinetic energy they have acquired since the previous collision; and this energy is transferred to thermal energy of the gas. So the gas warms up.

The situation, however, becomes different when the electric field is very strong. Then in the short time between collisions, some of the ions acquire sufficient kinetic energy to break up the gas molecules against which they collide. The new fragments are themselves ions which are accelerated by the electric field. They too become capable of disrupting the gas molecules which they encounter along their path. So an avalanche of more and more ions is produced. The gas suddenly acquires an enormous electrical conductivity, and a sudden discharge occurs. Geiger counters detect charged particles by using such an avalanche to turn the slight ionization along the track of a particle into an appreciable electric pulse.

B. Thermionic Emission. A somewhat similar relation between current and potential difference occurs for the electrons boiled out by a heated piece of metal (Fig. 27 – 11). When the potential difference between the heated filament and the surrounding collecting plate is big, we find a saturation current. Raising the potential difference does not change the number of electrons that cross from filament to plate, because all the electrons boiled out of the filament per second at a given temperature are already being pulled over to the plate. When the potential difference is low enough, however, the current drops beneath the saturation value. Apparently the electrons emitted from the filament do not move away sufficiently fast; and while they are in the neighborhood of the filament they push back electrons which would otherwise boil out. Detailed calculations actually show that this "space charge" limitation accounts for the current-voltage curves observed.

If the potential difference between filament and plate is reduced to zero, we observe something else of interest. A small current still passes. Even when the plate is made slightly negative with respect to the filament, a few electrons continue to get across. Apparently some electrons are boiled out of the filament with sufficient kinetic energy to go across despite a slight retarding potential difference. Careful



27–11 Current vs. potential difference curves for electrons boiled out of a heated wire at different filament temperatures in a diode.

measurements show that the potential difference necessary to stop the current is about two-tenths of a volt.* This corresponds to the thermal energy of a particle at the temperature of the filament.

C. Electrons in Metals. As a last example of the relation between current and voltage, we shall take a metal conductor. Then the relation between potential difference and current is a particularly simple one: the two quantities are proportional to each other. That is,

$$V = RI$$

The law expressed by this equation is called Ohm's law, and the proportionality constant R is called the *electric resistance* of the conductor. If V is measured in volts and I in amperes, R is measured in a unit called an ohm.

Ohm's law says that the current in a metal wire is proportional to the potential difference between one end and the other. Because the electric field in a given piece of metal is proportional to the potential difference imposed on its ends, we can just as well say that the current is proportional to the electric field. We can understand this law by returning for a moment to the model of the conductivity of metals, in which we picture electrons running freely around inside the metal like the molecules of a gas (Section 23 – 12). Because the electrons are so light, at normal temperatures they have tremendous speeds—about 10^5 meters per second. They rush around inside the metal, colliding with the positive metal ions with great frequency. The average time between collisions of a single electron with one or another of the metal ions is extremely short, only about 3×10^{-15} second. It is determined by the distance between metal ions and the *thermal speed* of the electrons. Furthermore, at each collision the electron bounces off in a new direction. So a single electron runs in all directions in its random thermal dance.

Now we add the electric field to our picture. In the short time between collisions, it accelerates the electrons along the wire. The small extra velocity which the electrons acquire along the wire is just proportional to the strength of the field and to the time between collisions. Because of the random changes in direction at each collision, or at worst after every few collisions, the extra velocity along the wire is wiped out and has to be reestablished by the action of the field. Our complete picture then looks like this: the electrons rush around in all directions at huge speeds, but superposed on top of these random motions is a small systematic velocity along the wire. This systematic velocity is proportional to the strength of the field and to the time between collisions. With a potential difference of a few volts across a meter of wire, the systematic velocity is about a centimeter a second. It is this systematic velocity that results in a current. The bigger the systematic velocity the bigger the current; so the current is proportional to the electric field. Our model therefore predicts the experimental result embodied in Ohm's law.

The model also makes another prediction about the electrical resistance. It should increase with increasing temperature. As the temperature goes up, so does the speed of the electrons. Therefore the time between collisions diminishes; and the same field yields less current. This conclusion agrees with the experimental observation that the resistance usually does become greater. This agreement is not perfect, however; detailed experiments show that the resistance usually increases in direct proportion to the temperature, while the model predicts that the resistance should be proportional to only its square root. The discrepancy is explained by two facts. In the first place, because of its high density, we should not expect the electron gas to behave like an ideal gas. The electrons are so close together that the forces between them always influence their motion. Secondly, when a gas of identical particles is dense, quantum effects, which we shall study later in this book, become important. When these effects are included, the improved model is in accurate correspondence with the experimental facts. Our model remains useful as one of the essential steps to a more complete description.

---

*To make these measurements it is necessary for all the hot metal to be at the same potential. Otherwise the potential differences from one place to another in the metal may be as large as the potential difference we wish to measure. We therefore use a separate heater instead of having a large electric current flowing in the hot metal.

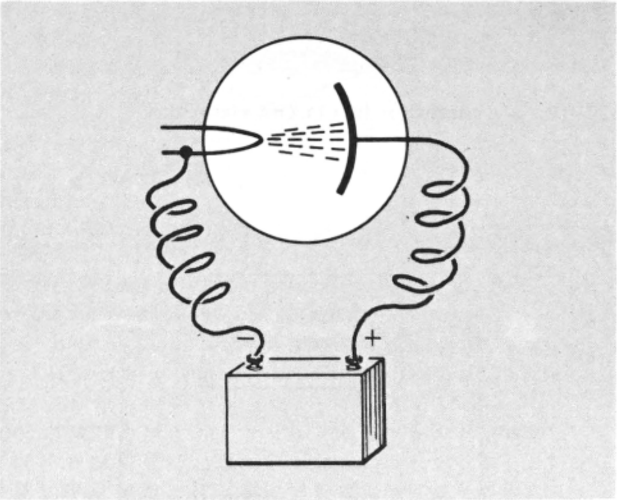
27–5 AN OVERALL VIEW OF A CIRCUIT

When a battery lights a flashlight bulb, runs an electric motor, or maintains the current in an electrolytic cell, the battery delivers energy to the external circuit (the bulb, the motor, or the electrolytic cell). There this energy is transformed into heat, or into mechanical energy. If the circuit contains a radio tube with a heated filament, the electrons gain a large amount of kinetic energy as they cross the vacuum from the filament and crash into the plate. In a tube designed so that the electrons hit a small region of the plate, they may make it white hot. In a metal wire the energy supplied by the battery also goes initially into kinetic energy of motion of the electrons. Then as the electrons collide with the atoms of the wire this kinetic energy is transformed into the energy of random motion of the atoms. Because the collisions take place frequently, the electrons never acquire much kinetic energy. The energy fed in becomes heat almost at once. If the circuit contains a motor, most of the energy may be used to raise weights, though some of the energy always comes out as heat.

Let us think of the circuit sketched in Fig. 27 – 12. Each electron passing from filament to plate acquires a kinetic energy proportional to the potential difference maintained between them by the battery. Each electron that goes across takes one negative charge from the filament and cancels one positive charge at the plate. If you disconnected the battery by opening the switch, the current would soon stop. With the battery connected, when one electron moves across, another one is forced along the wire from the negative battery terminal, taking the place of the one that has left. A current equal to the current from filament to plate is thus flowing through the wire. A small electric field in the wire forces the electrons to move. So when the current flows, the different parts of the conducting wire are not exactly at the same potential. If the wire is long and thin enough, and if the current is sufficiently great, the potential difference from one end to the other will be big enough for us to measure.

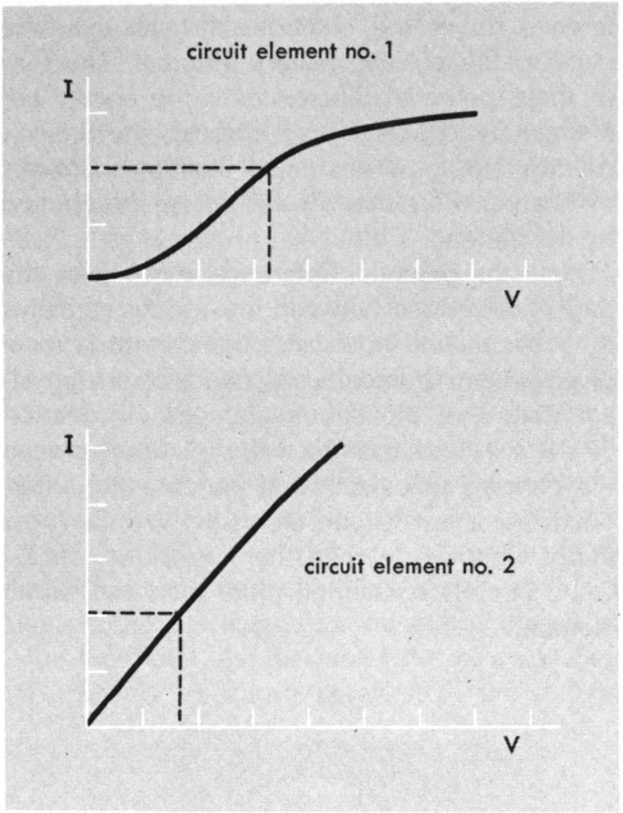
The potential difference between the battery terminals—the energy per elementary charge supplied by the battery—is split up into potential differences across the various elements of the circuit. A small potential difference drives the electrons from the negative battery terminal to the filament. It accounts for the energy given to each electron and then dissipated in heat as the electron moves through the wire. A large potential difference accelerates the electrons from the filament to the plate and accounts for the kinetic energy with which they crash into the plate. Finally, another small potential difference drives the electrons through the wire back to the positive battery terminal. The sum of these potential differences is the energy per elementary charge moved through the battery. All this energy is dissipated, heating the wires or heating the plate. All this energy is supplied by the battery.

How the potential difference imposed by the battery is divided between the various elements of a circuit and how large the current is for a given potential difference depends on the physical nature of the circuit elements. In Section 27 – 4 we have examined the relation between the currents that occur and the potential differences for a few circuit elements. We can now apply what we learned there to more complicated circuits containing more than one circuit element.

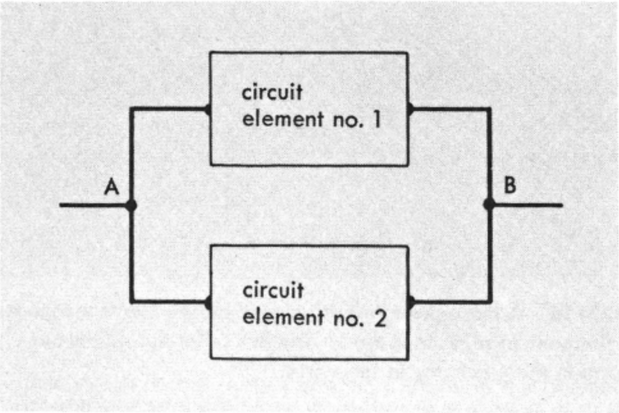


27–12 A large potential difference moves electrons from filament to plate in a diode. Smaller potential differences move the electrons in the wires.

Suppose that we have two circuit elements for which the curves of current vs. potential difference are known, for example, the circuit elements represented by the graphs of Fig. 27 – 13. If these two elements are connected in parallel as in Fig. 27 – 14, the potential difference V across their ends from A to B is automatically the same for both elements. The current that comes in at A and goes out at B splits up, and part of it passes through each circuit element. The total current is the sum of the currents through the two elements. It can be found by adding the current I_1 through element 1 at the particular value for the potential difference and the current I_2 through element 2 at the same value of the potential difference. By adding up the two currents determined from Fig. 27 – 13 at each value of potential difference, we can make a graph of the total current $I = I_1 + I_2$ passing from A to B versus the potential difference V (Fig. 27 – 15). In this way we combine the two elements in parallel into a new single element represented by the new I versus V graph. This procedure can be extended to handle more circuit elements in parallel by adding the individual currents at each potential difference. Thus, we can always replace a group of circuit elements connected in parallel by an equivalent circuit element that acts in the same way as the whole parallel combination.



27–13 I-V curves for two circuit elements.



27–14 Two circuit elements connected in parallel.

For any potential difference V across the parallel combination the energy per elementary charge is still given by the potential difference. Consequently, the power used in the parallel combination is IV where I is the total current—the current through the combination viewed as a single element. This power used is made up of the power used in each of the individual elements. For example, if there are two elements,

$$IV = I_1V + I_2V$$

where I_1 and I_2 are the currents through the individual elements that we have connected in parallel. In this equation, each of the terms on the right is the power used in a particular individual circuit element, and their sum on the left is the total power consumed by the whole group combined in parallel.

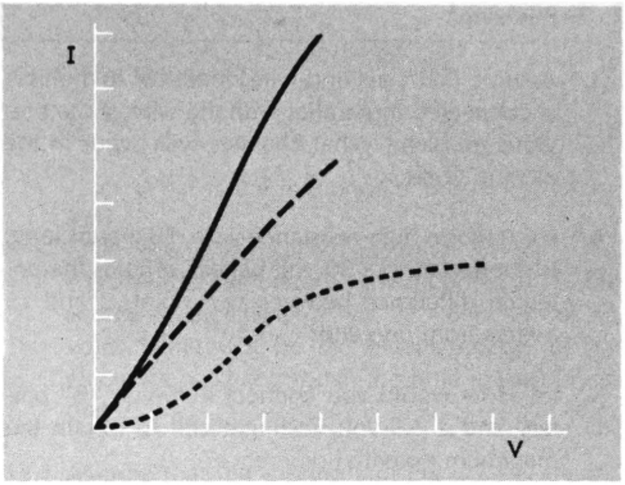
If we connect individual circuit elements in series, the analysis is somewhat different. The current must be the same in each of the individual elements. Let us start by assuming that this current has a particular value. By using the graphs of current vs. potential difference for each individual circuit element, we can read off the corresponding potential differences. Then to find the potential difference from one end of the series circuit to the other, we must add all of these individual potential drops. On Fig. 27 – 13, we have marked the same current on each of the current vs. potential difference curves and indicated by the dashed lines how to find the corresponding potential drops. When these two elements are in series, the potential drop across the series circuit is given by the sum of the two potential drops we read on the horizontal axes of the individual *I* versus *V* graphs. By following this scheme for each possible current, we can establish the relation between the current and the potential difference across the whole series circuit. We can again graph it as a curve of current vs. potential difference (Fig. 27 – 16). In this way we can turn the series circuit into an equivalent single circuit element represented by this new curve of current vs. potential difference.

Now that we have found the behavior of this equivalent single circuit element, we can find the current through the series circuit when any specified potential difference is applied across its ends. As always, this potential difference represents the energy used per elementary charge passing through the circuit. Again the power consumed is IV, where V is the potential difference from one end to the other of the whole series circuit. In this case, however, V is the sum of the individual potential drops across the individual elements forming the series circuit. If the circuit is made of two elements, that sum is $V = V_1 + V_2$. Consequently,

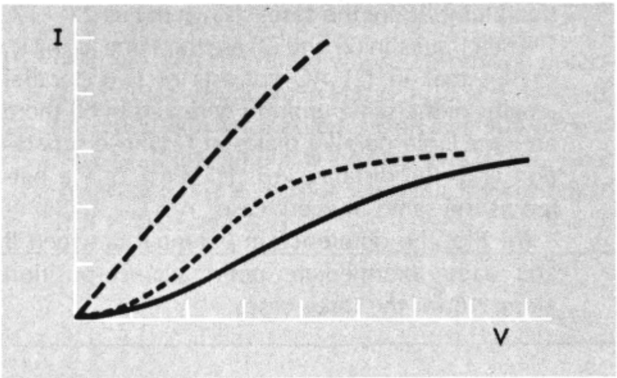
$$IV = IV_1 + IV_2$$

In this equation the power consumption of each element is represented by a term on the right-hand side; and the total power consumption IV is just the sum of the power used in each individual element.

We have seen how the potential differences and the currents actually come into adjustment in a real circuit. The results fit into our overall view of a circuit. In every case the sum of the potential difference taken in series around the circuit matches up with the EMF's of the batteries or generators supplying the energy per elementary charge. The energy passed out per elementary charge is equal in each circuit element to the difference in potential from one end of the element to the other. The energy thus transferred into other forms such as chemical stored energy, mechanical energy or heat is supplied by the applied EMF; and the conservation of energy shows us that the sum of the potential differences is always equal to that applied EMF.



27–15 I-V curve for the equivalent circuit element made by placing the circuit elements of Fig. 27–13 in parallel. It is obtained by adding the currents of the two individual circuit elements when the potential difference has the same value for both.



27–16 I-V curve for the equivalent circuit element when the circuit elements of Fig. 27–13 are connected in series. It is obtained by adding the potential differences across the two circuit elements when the current has the same value through both. You can visualize this addition by looking at the picture from the left. There if you pick a given current, the potential difference needed to produce it in circuit element 1 is given by the distance along the V axis to the dotted curve. The potential difference needed to produce this current in circuit element 2 is given by the distance along the V axis to the dashed curve. The sum of these two potentials is the potential difference needed across the circuit elements in series to produce the same current.

QUESTIONS

1.* What can you conclude about the field in a current-carrying wire from the fact that the current is always in the direction of the wire?

2.* A uniform high-resistance wire, 10 meters long, is connected to a battery having an EMF of 30 volts as in Fig. 27 – 3. What is the electric field in this wire?

3.* Assume that a second wire, identical to the first, is connected in parallel with the wire of the previous problem. What changes will occur in the electric field?

4.* If a uniform high-resistance wire, 20 meters long, is connected to a 30-volt battery, what is the potential difference between two points 3 and 15 meters from one end?

5. (a) How would you connect a 90-volt "B" battery and a 1.5-volt flashlight cell to obtain the maximum potential difference?

(b) How big is that potential difference?

(c) Suppose you disconnect the 90-volt battery, turn it end for end, and then reconnect it. What potential difference do you get?

6. (a) Plot the electric field intensity against position along *AB* for the cases shown in Fig. 27 – 17.

The circuits in (2) and (3) are the same as in (1), except that in (2) *AC* consists of two parallel lengths of the same uniform wire and in (3) there are two large parallel plates at C whose separation is $\frac{1}{5}$ the distance *AB*. The EMF of the battery is the same in every case.

(b) Plot the difference in potential between *B* and each intermediate point against position along *AB* for the three cases.

7. Two uncharged parallel plates are connected as shown in Fig. 27 – 18. When the switches are closed, connecting the plates to a 270-volt battery, charges flow until each plate has 6.4×10^{-9} coulomb.

(a) How much work does the battery do in charging the plates?

(b) How much energy is stored in the charged plates?

(c) Where does the rest of the energy go?

8. The two plates of Question 7 are charged with the 270-volt battery so that each plate has 6.4×10^{-9} coulomb or 4.0×10^{10} elementary charges. The battery is disconnected and the negative plate is heated until electrons are emitted from its surface and accelerated toward the positive plate.

(a) What is the kinetic energy in joules of an electron arriving at the positive plate when this process has just begun?

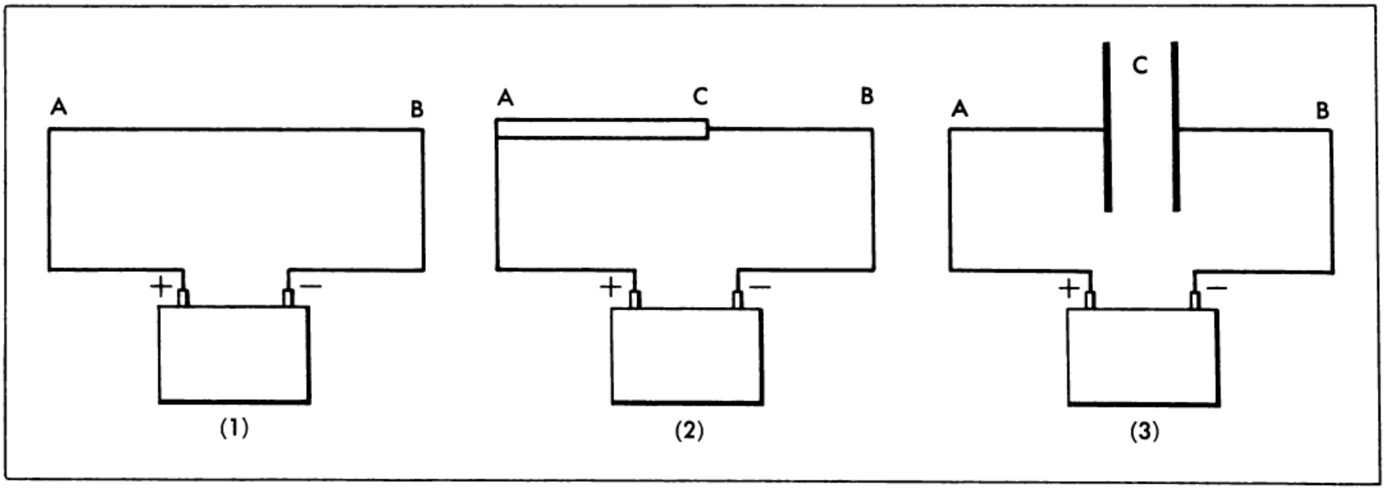
(b) What is the kinetic energy of one of these electrons when 3.0×10^{10} elementary charges are left on the negative plate?

(c) What is the total kinetic energy which these electrons bring to the positive plate if the process continues until the plates are completely discharged?

9.* Why is it desirable that an electrometer should not remove much charge from the object on which it takes a reading?

10.* Under what condition can you use a current-type voltmeter to measure potential difference between two points in a closed circuit?

11. Suppose you were given two electrometers which measure potential differences in the 10,000-volt range. One of them is calibrated but the other



27–17 **For Question 6.**

is not. You are also given a glass rod and a piece of cloth, but you are forbidden to buy any batteries. Can you calibrate the second electrometer by connecting it to the first?

12. An electron gun shoots a stream of electrons between two deflecting plates of length $l = 6.0$ cm (Fig. 27 – 19). The electron mass is 0.91×10^{-30} kg. The potential difference between the filament and plate of the gun is 2500 volts ($2.5 \times 10^3 \times 1.6 \times 10^{-19}$ joule per elementary charge). The electric field E between the deflecting plates is 1000 newtons per coulomb (1000 newtons per 6.25×10^{18} elementary charges).

(a) If the deflecting plates are separated by 0.20 cm, what is the potential difference from one deflecting plate to the other?

(b) What is the speed of the electrons leaving the gun?

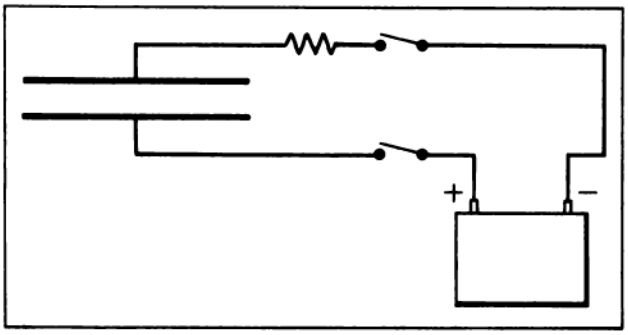
(c) How long do the electrons take to pass through the length l?

(d) As an electron passes between the deflecting plates, it picks up a sidewise velocity because of the electric force exerted on it. How big is this sidewise velocity when the electron leaves the deflecting plates?

(e) How far is the electron pushed sidewise in passing between the plates?

13. (a) For the apparatus used in Question 12, find a formula for the sidewise displacement of any ion accelerated through the same potential difference V in the gun and deflected by the same deflecting field E. Write the formula in terms of E, V, and l, the length of the deflecting field.

(b) What is the influence of the mass of the ion?



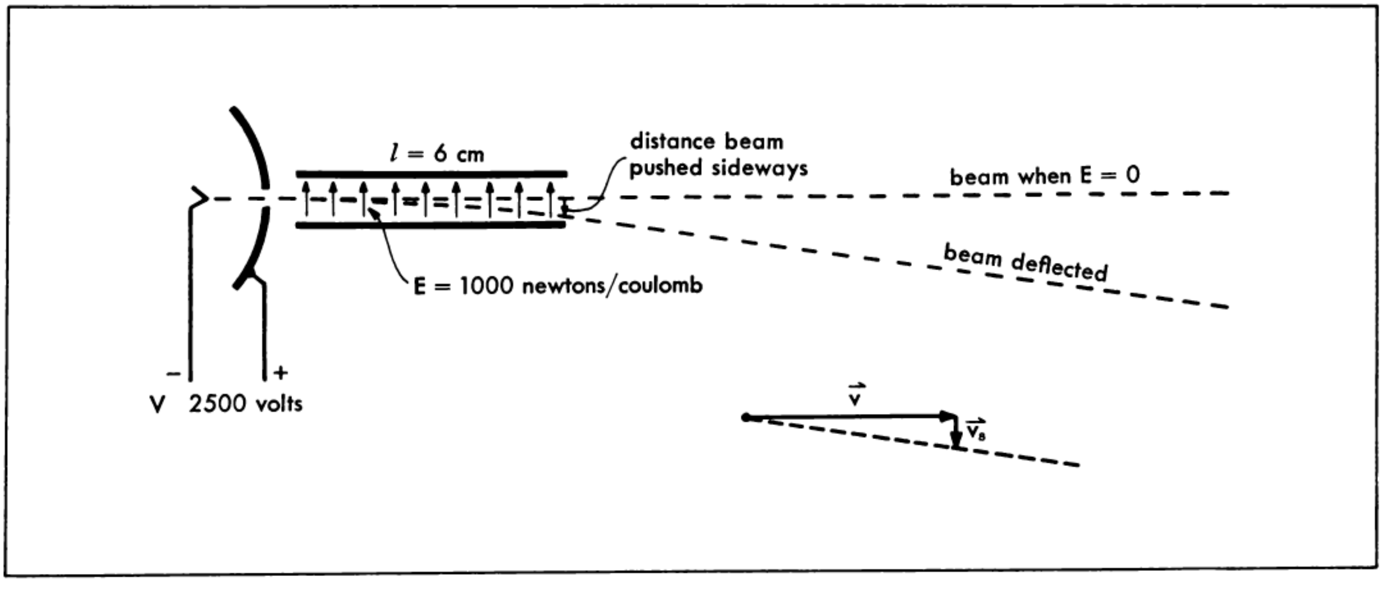
27–18 For Question 7.

(c) What is the influence of the charge of the ion?

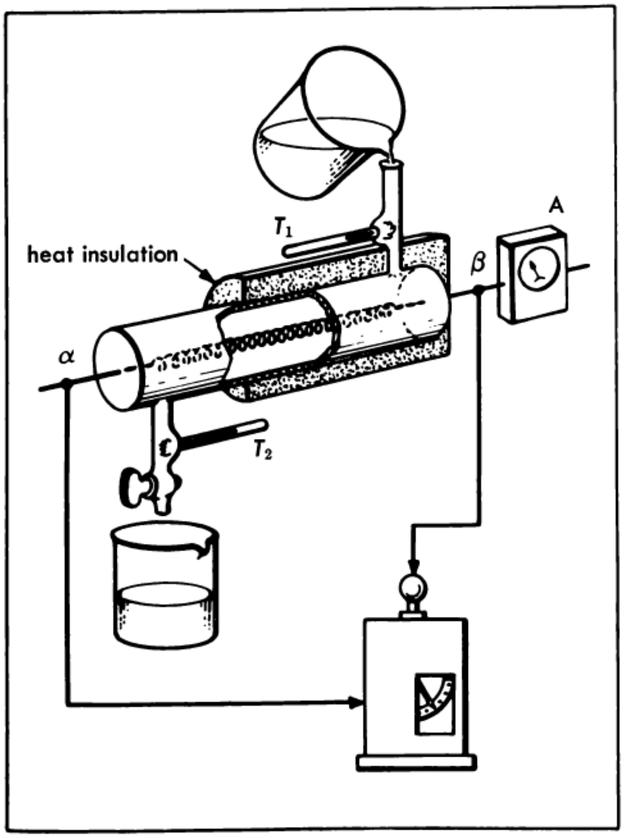
14.* Assume that the battery shown connected to the source of electrons in Fig. 27 – 8 is one of 45 volts. What will be the kinetic energy of the electrons passing through the hole in the curved plate?

15.* Assume that the electrons in Fig. 27 – 8 are accelerated by a 45-volt battery and that the potential divider and two batteries across the parallel plates are replaced by a 30-volt storage battery. Describe what happens to the kinetic energy of an electron which passes through the hole in the plate at the left and strikes the plate at the right.

16.* Assume that the diode whose current vs. potential curves are shown in Fig. 27 – 11 is connected, in series with a variable resistor, to a 90-volt battery. If the filament is heated to 2100°K, what resistance in the resistor will limit the current to 1.2×10^{-3} ampere?



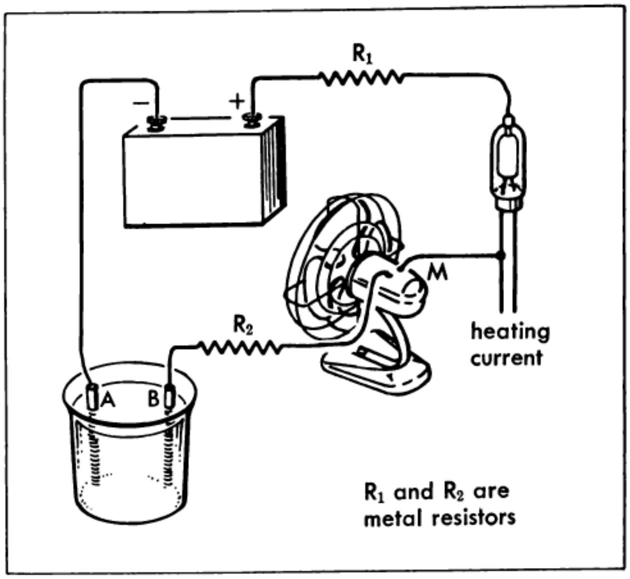
27–19 For Question 12.



27–20 For Question 17.

17. Figure 27 – 20 shows a device for calibrating the ammeter A. It is an insulated can with a coil of wire in it. Water flows in at the top of the diagram and passes thermometer T_1. Then, after flowing past the coil of wire, it flows out at the bottom of the diagram past thermometer T_2. The difference of potential across the coil in the calorimeter is measured, using a calibrated electrometer, so that we know the potential difference between the points α and β. In the experiment (i) the flow of water is 1 gram per second, (ii) thermometer T_2 reads a temperature 1°C higher than thermometer T_1 and (iii) the electrometer reads a potential difference of 2.09 volts.
    (a) What is the current through the ammeter?
    (b) How would you calibrate an ammeter with this device?

18. In the circuit shown in Fig. 27 – 21, two resistors, a diode, an electric motor M, and an electrolytic cell are connected in series with a battery. The difference of potential across each of these circuit elements is measured and it is found that they are all the same. A certain amount of heat is dissipated per second in the first resistor. Compare this with the amount of heat dissipated per second:
    (a) In the second resistor.
    (b) In the diode.
    (c) In the motor which runs a fan.
    (d) If there is actually less heat dissipated in the electrolytic cell than in resistor R_1, which terminal, A or B, would be positive when the cell is used as a source of potential difference?



27–21 For Question 18.

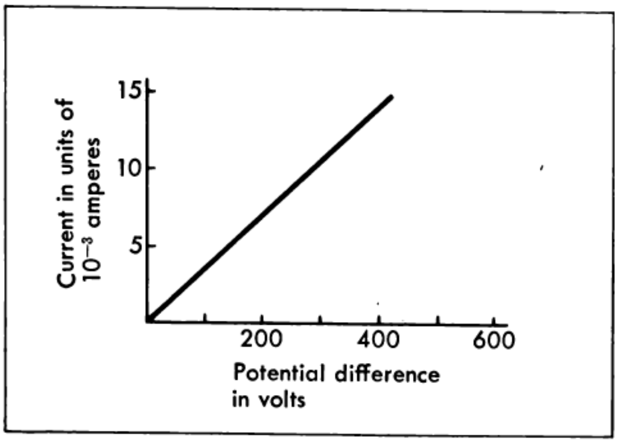
19. (a) In Question 17, what is the resistance in ohms of the coil?
    (b) Show that the power dissipation from the coil is given by I^2R.

20. If the conduction electrons in a metal move freely through it like the molecules of a gas, what is the mean thermal speed of the electrons at 27°C?

21. (a) A copper wire has n conduction electrons per unit volume. The wire is carrying a current I produced by a systematic drift of the conduction electrons parallel to its length with an average velocity v. The area of cross section of the wire is A. How many electrons pass a marked point on the wire during the time Δt?
    (b) If each electron carries a charge q, what is the current in the wire?
    (c) Suppose that each copper atom in the wire contributes one conduction electron. Calculate the number of conduction electrons per cubic centimeter, taking the atomic mass of copper as 63 and its density as 8.9 gm/cm^3.
    (d) A common size of wire has an area of cross section of 0.83 mm^2. A reasonable current for this wire is 5 amperes. What is the systematic drift velocity of the electrons along it?

22.* If the marks in Fig. 27 – 13 refer to volts and amperes and we connect the two circuit elements in series with a battery, what voltage would be needed to give a current of 2.0 amperes?

23. We have two lengths of wire made of exactly the same metal and of exactly the same cross-sectional area. One wire is twice as long as the other.



27–22(a) For Question 25.

(a) When the same difference of potential is imposed on the ends of both wires, what is the ratio of the currents that will pass through the wires?

(b) How does the resistance of a wire depend on its length?

24. There are two wires of identical material of exactly the same length and the same cross-sectional area.

(a) If the same difference of potential is imposed across their ends, what is the ratio of the current that will pass through them?

(b) If the wires side by side are viewed as a single conductor, what is the ratio of the current through this combined conductor to the current through one of the single wires with the same potential difference across its ends?

(c) If we have a wire of twice the cross-sectional area and the same length and material as each of the original ones, what current do you expect will flow through it compared to one of the original wires, with the same potential difference across its ends?

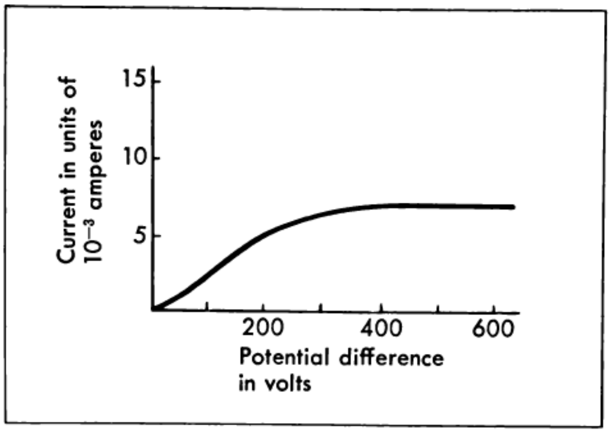
(d) What do you expect to be the dependence of the resistance of a wire on its cross-sectional area?

25. In Fig. 27 – 22 (a), the current through a metal circuit is plotted as a function of the potential difference across its ends. In Fig. 27 – 22 (b), we plot the current through a diode as a function of the potential difference across it.

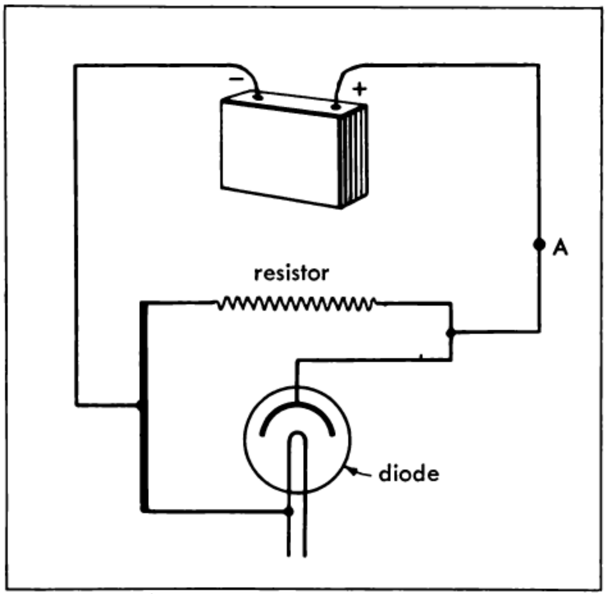
(a) If a current of 5×10^{-3} ampere passes through both circuit elements in series, what is the potential difference from one end to the other of the resistor? What is the potential difference across the diode?

(b) If the diode and the resistor are connected in parallel as in Fig. 27 – 22 (c), and the potential difference across the diode is 100 volts, what is the current through each of the circuit elements?

(c) What then is the total current passing point *A* in the figure?



27–22(b) For Question 25.



27–22(c) For Question 25.

(d) What is the resistance in ohms of the resistor?

26. The same diode and resistor as were used in Question 25 are connected in series. The potential difference from one end of the circuit to the other is 1000 volts.

(a) What is the current in the circuit?

(b) What is the potential difference between one end and the other of the resistor?

27. Connect a 270-volt battery in series with a 1,000,000-ohm and a 2,000,000-ohm resistor.

(a) What potential difference do you expect across the megohm resistor?

(b) What potential difference do you expect across the 2,000,000-ohm resistor?

Now suppose you measure the potential differences with a voltmeter.

(c) If the voltmeter has a resistance of 10 megohms, what potential difference would you expect across the 2,000,000-ohm resistor?

(d) What potential difference would you expect to read across the megohm resistor?

28

The Magnetic Field

WHEN Albert Einstein wrote his life story at the age of 67, he recalled the day when he was a little boy of four, happy to get a new toy from his father. That toy was a compass needle; and the wonder it aroused remained with Einstein all his life. Most of us have shared that wonder, and a good many of us as children experienced the fascination of a horseshoe magnet, which attracts iron. In this chapter we shall discuss magnetic forces and the concept of the magnetic field by which we can best interpret them.

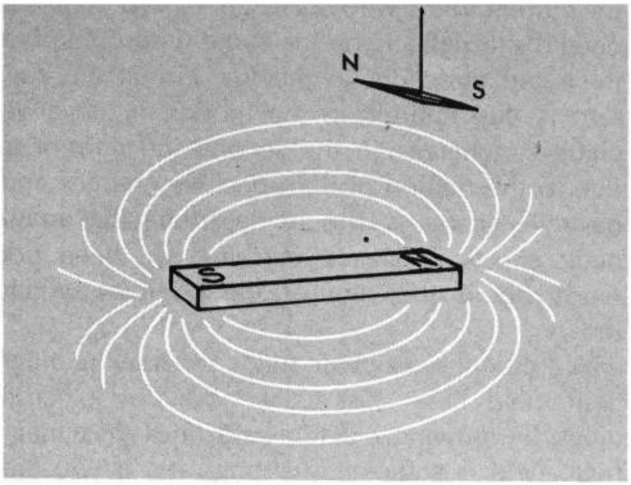
28–1 THE MAGNETIC NEEDLE

A compass needle free to rotate about a vertical axis will line up in the north-south direction. We say that it aligns in the magnetic field of the earth (which originates deep in the core of the globe). We call the tip of the needle that points north the northern tip.* In any region where a compass needle feels forces that tend to align it, we say that there is a *magnetic field* (Fig. 28 – 1). The direction of the field is the direction taken by a compass needle which is free to line up. We say that the field points in the direction that runs from the southern to the northern compass tip. Just as we used grass seeds to show the direction of an electric field, so we use a magnetic needle to explore the direction of a magnetic field. You can see that a compass needle itself has a magnetic field by bringing up another compass needle and watching the two influence each other: each pulls on the other. Each produces a magnetic field, and tends to line up in the field of the other.

As early as 600 B.C. the ancient Greeks were acquainted with magnetism. Thales of Miletus, often called the father of Greek science, knew about a mineral (magnetite) which attracts ordinary iron, and found

---

*Actually, the northern tip of the needle points *approximately* to the geographic North Pole as long as we are south of the Arctic Circle (and north of the Antarctic Circle). In the Northern Hemisphere the northern tip of a compass needle pivoted on a horizontal axis also points down, measuring the "dip" of the earth's magnetic field.



28–1 A compass needle in a magnetic field lines up in the direction of the field.

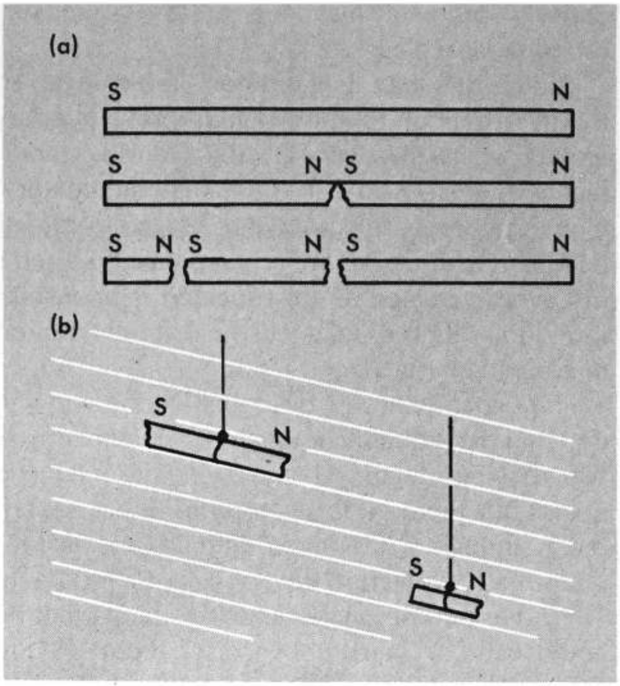
that the iron itself becomes magnetized by touching the magnetic mineral. The Chinese discovered, probably about the eleventh century, that a magnet acts as a compass. The unifying idea that the earth itself is a magnet was provided by William Gilbert, who was working at the court of Queen Elizabeth of England in 1600, just about the time Shakespeare was writing *Hamlet.*

One of the most striking features of magnets is that if we break one in two, each piece is also a magnet (Fig. 28 – 2). We might ask how far we can continue subdividing before "half a magnet" is no longer a magnet. The answer is that this never happens. We can continue right down to the subatomic scale and find that electrons, protons, and neutrons themselves are magnets.

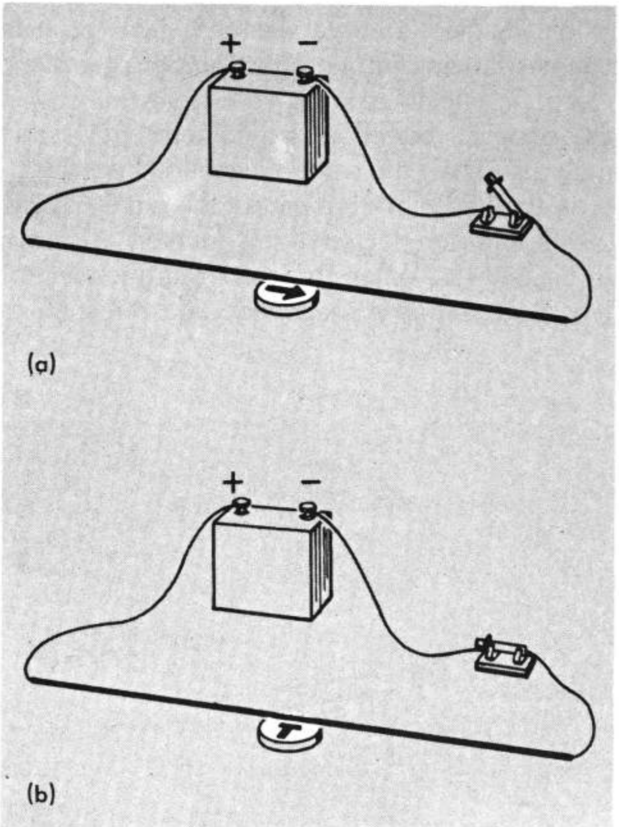
28–2 MAGNETIC FIELDS OF MAGNETS AND CURRENTS

Magnetite and magnetized iron are not the only sources of magnetic fields. Let us perform an experiment in which we produce a magnetic field without using such materials.

We attach a long piece of wire to the terminals of a battery through a switch as shown in Fig. 28 – 3. With the switch open, we hold the wire above a compass needle and parallel to it. Then we close the switch. If the current in the wire is sufficiently strong, we see the compass needle suddenly deflected. It now points across the wire. We therefore conclude that electric



28–2 (a) When we break a magnet, each piece has opposite poles at its ends. (b) Any of these pieces will line up in a magnetic field.



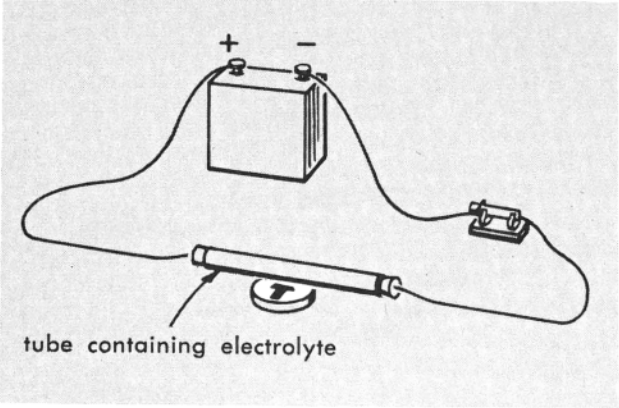
28–3 (a) A wire is placed over a compass needle and parallel to it. The switch is open, and no current flows in the wire. (b) When current flows, the needle is deflected and points across the wire.

currents produce magnetic fields in the surrounding space.

Today this fact is common knowledge, so that it is difficult to appreciate the revolutionary impact of its discovery by the Danish schoolteacher Oersted early in the nineteenth century. Until that time, electric and magnetic effects had been regarded as entirely separate. Oersted's discovery revealed an unsuspected relationship: it tied the origin of magnetic fields to the motion of electric charges.

The experiment last described is one you can do, and is probably enough to convince you that moving electric charges generate magnetic fields. But there is an even more direct experiment, though a far more difficult one. It was performed by Henry Rowland in Baltimore in 1876. He put the biggest electric charge that he could onto a hard-rubber disc about 20 cm across. Then he spun the disc at about 60 revolutions per second. In this way he could look for the magnetic effect of the moving charges directly. He found that the spinning electric charge did produce a weak magnetic field.

In Chapter 25 we learned that positive charges moving in one direction have the same electrical effects as negative charges moving in the opposite direction. Experiment shows that they also have the same magnetic effects. Suppose that we replace a part of the circuit of Fig. 28 – 3 with an electrolytic conductor, as shown in Fig. 28 – 4. When the current in the electrolyte is the same in magnitude and direction as the current in the wire of Fig. 28 – 3, the compass needle still deflects by the same amount and in the same direction. In the first circuit the current is due entirely to the motion of negative charges, whereas in the second the current is due to the motion of both positive charges and negative charges. Thus in describing the magnetic effects of steady currents, we need not worry about the sign of the particles which move.



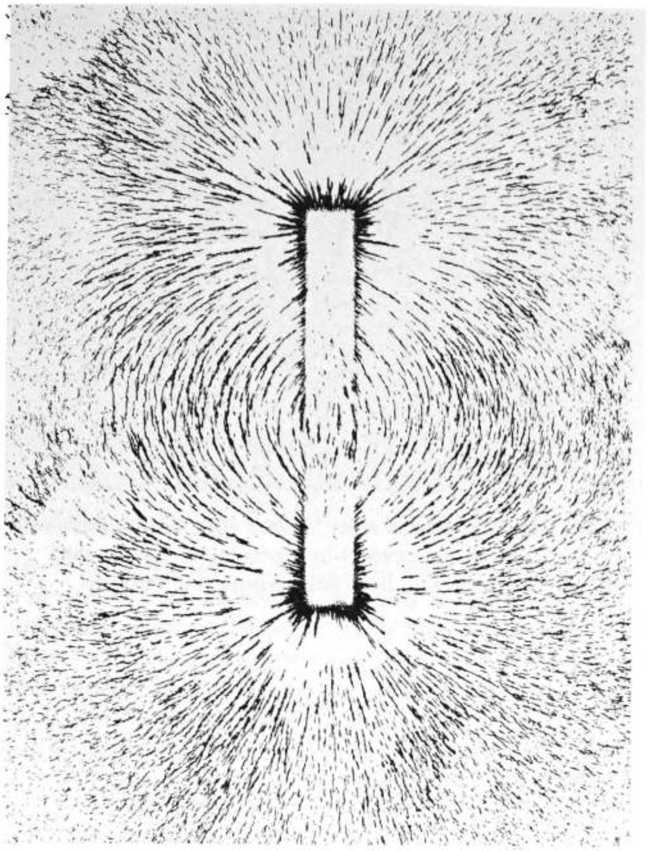
28–4 If we replace part of the circuit of Fig. 28–3 with a tube containing electrolyte, the compass needle deflects just as before if the current in the circuit is the same as before.

We describe magnetic fields by drawing magnetic field lines, just as we describe electric fields by drawing electric field lines. If a magnetic field is sufficiently strong, and if you are not interested in great accuracy, you can "see" the field lines by spreading some fine iron filings on a piece of paper. The grains of iron line up along field lines like grass seeds in an electric field (see Section 24 – 3). Tap the paper gently if the filings stick. The photographs in Fig. 28 – 5 are iron-filing pictures.

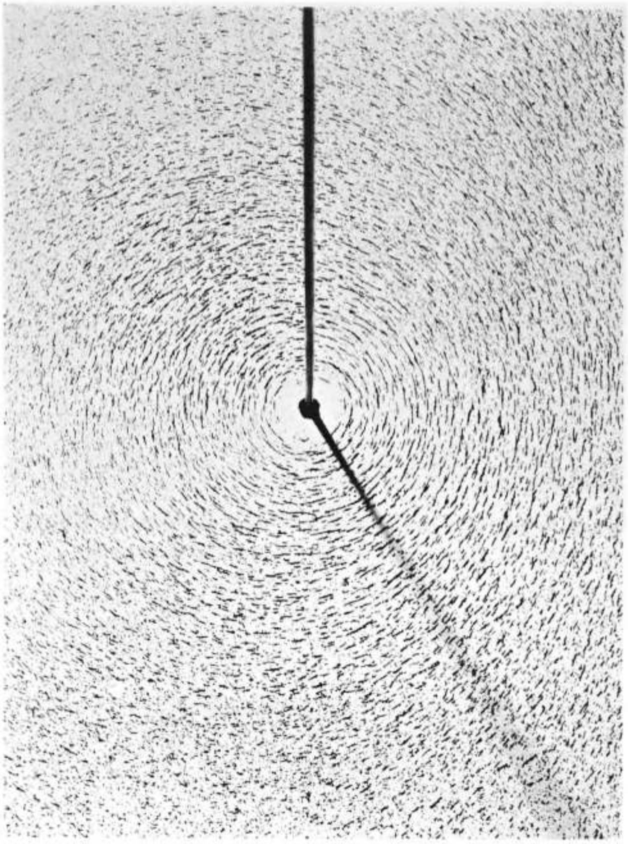
Of course, we do not get the direction of the field, only the pattern of the field from the iron filings. We can determine the direction by placing a small compass needle at a number of different positions in the field. Then we can draw the field lines for the various patterns, indicating their directions with arrowheads. Figures 28 – 6 and 28 – 7 show the magnetic fields produced by various magnets and currents. You may compare them with Figs. 24 – 8 and 24 – 10 showing various electric fields.

Electric field lines from charges at rest begin and end at the charges producing the field; but the lines of the magnetic fields produced by currents have no beginning nor end. They encircle the wires carrying the current (Fig. 28 – 7). The lines of the magnetic fields produced by permanent magnets (Fig. 28 – 6) *appear* to begin and end at the surface of the magnets, but this is only because we have not drawn the lines inside the magnets where we cannot place an ordinary compass needle. Using neutrons as the compass needles, we can explore the inside of a permanent magnet. We then see that the lines do not stop at the surface.

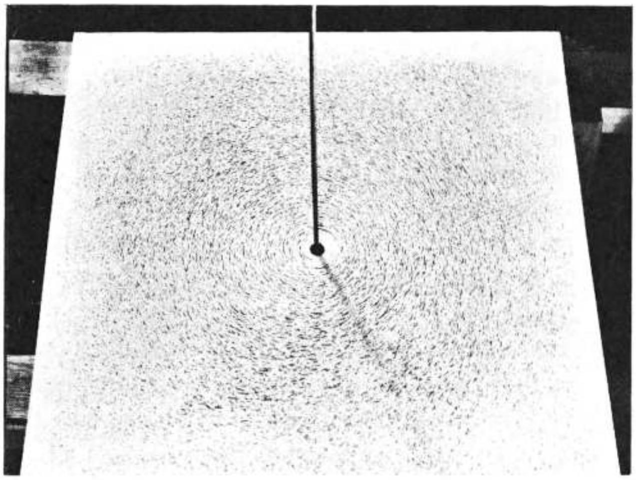
You can remember how the direction of the field is related to the direction of the current by the "right-hand" rule. In imagination place your *right* hand with the thumb in the direction of the current; then your fingers curled around



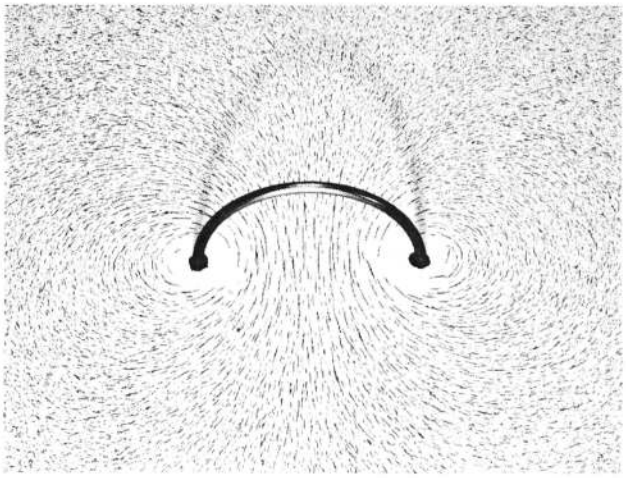
(a)



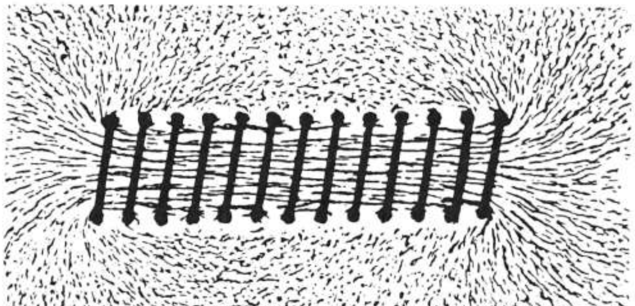
(b)



(c)

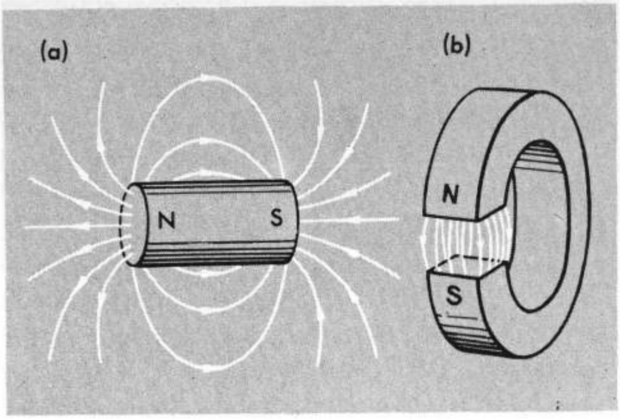


(d)

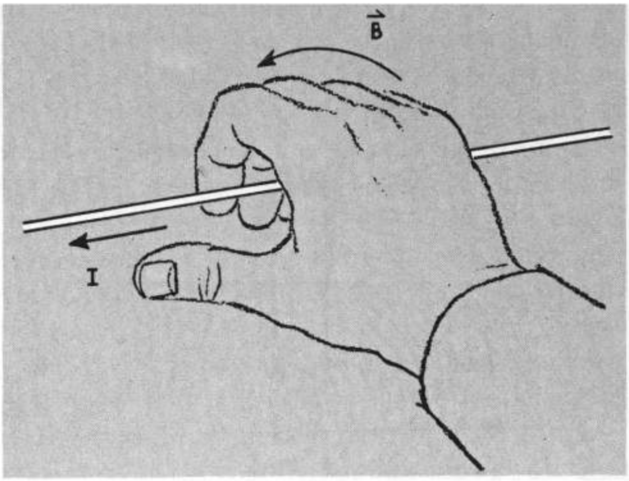


(e)

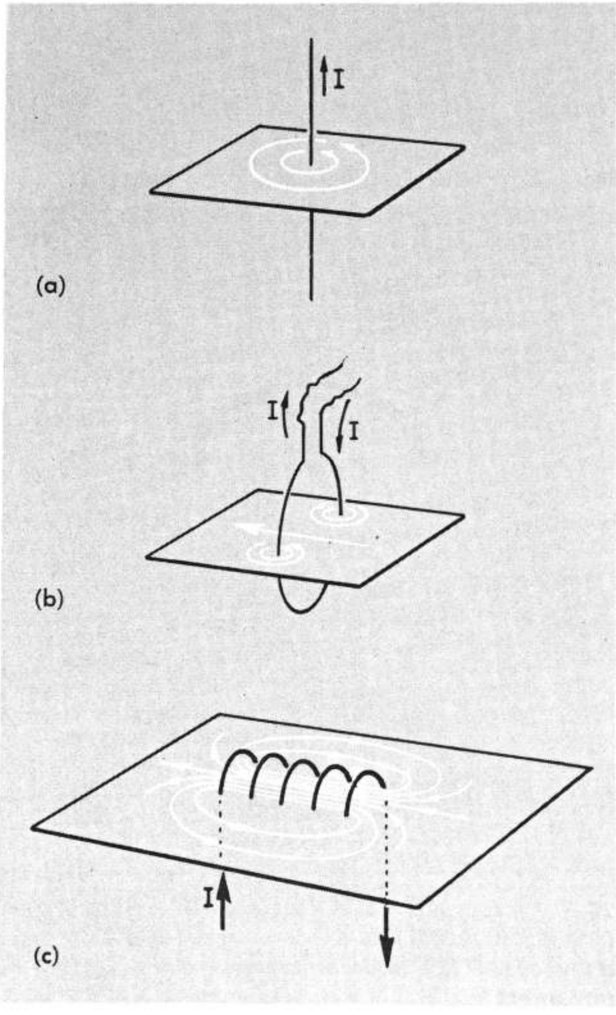
28–5 Photographs of iron-filing magnetic field patterns. (a) A bar magnet. (b) Current in a long straight wire. (c) An overall view of the arrangement used to get (b). (d) Current in a loop of wire. (e) Current in a solenoid. [(e) From "Textbook of Physics," R. Kronig, Ed., Pergamon Press, Inc., 1959.]



28–6 Diagrams of the magnetic field lines of permanent magnets. (a) A bar magnet. (b) A horseshoe magnet.



28–8 The "right-hand rule." When your thumb points in the direction of the current in a wire, your fingers curl the way the magnetic field goes around the current.



28–7 Diagrams of the magnetic field lines around current-carrying wires. (a) A long straight wire. (b) A loop of wire. (c) A solenoid.

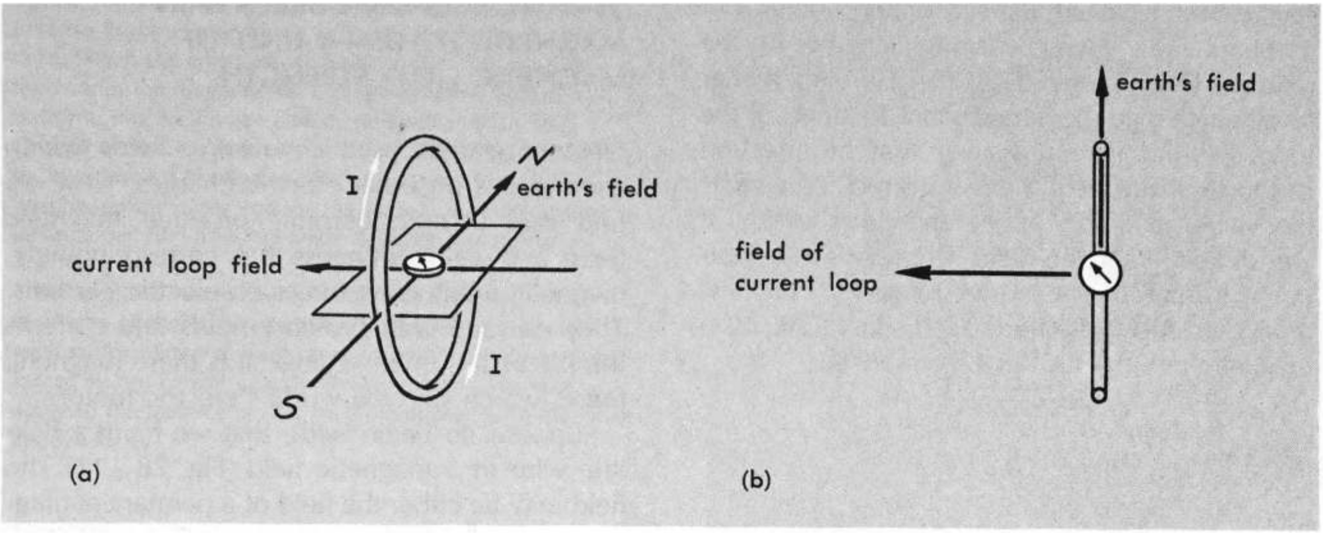
the wire point in the direction of the field (Fig. 28 – 8). Notice that this rule also gives the direction of the field through a loop of wire.

28–3 THE VECTOR ADDITION OF MAGNETIC FIELDS

We have seen that a freely swinging compass needle aligns itself in certain directions as it is moved around in a magnetic field, and we have used the direction in which it points to define the direction of a magnetic field. Furthermore, we have a good idea of what we might mean by the magnitude of a field, or magnetic field strength: with a compass needle close to a large magnet, it takes more force to twist the needle out of the direction of the field than with the compass in only a relatively weak field such as the field of the earth. If a magnetic field can be characterized by both a magnitude and a direction, it is natural to suppose that two magnetic fields, acting together, would add like vectors. Let us check up on this experimentally.*

---

*Not everything with magnitude and direction is a vector. For example, the rotations of a body in three dimensions cannot be added vectorially. Try rotating a book 90° around a horizontal axis and then rotating it 90° around a vertical axis. Now try doing the rotation in the opposite order. You do not get the same final result. The rotations do not add vectorially. (We know that vectors can be added in any order.)



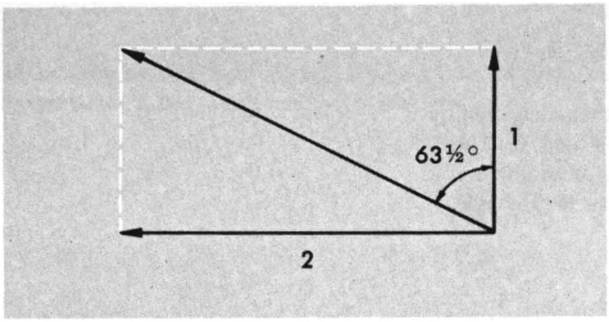
28–9 (a) A magnetic field, equal in magnitude to the earth's field (as measured by a horizontal compass needle), is added at right angles to the earth's field. (b) The horizontal needle points at 45° away from magnetic north. For any other angles between the equal-magnitude fields, the compass needle always bisects the angle. Magnetic fields appear to add like vectors.

We begin by investigating a simple case, the addition of two fields of equal magnitude. We shall use a small compass needle, pivoted on a vertical axis, to detect the resultant field. Since the earth's field is always present, let us choose this as one of our two test fields. The other test field can be produced with any convenient current arrangement for which we know the direction of the field. We shall choose a circular loop, because the field at its center is fairly uniform and has the direction of the axis of the loop [Fig. 28 – 5 (d)]. To make the field at the center of the circular loop equal in magnitude to the earth's field, we place the compass needle at the center of the loop and, when the current is zero, align the axis of the loop along the direction of the needle. We choose the direction of the current by the right-hand rule so that its field is opposite to the earth's field. As we increase the current in the loop, we find a value at which the compass needle swings freely. There is no force to align it in any particular direction; the needle indicates zero magnetic field. Thus the magnetic field of the earth has been just balanced by an equal and opposite magnetic field due to the current in the loop.

Now let us add these two equal fields together at right angles to each other. We rotate the loop about a vertical diameter so that its axis turns through 90°. If the current has not changed, we find that the compass needle now points at 45° to the direction of the earth's field (Fig. 28 – 9). This is what we expect if magnetic fields add like vectors. But one experimental check is not enough. We can try other angles between the directions of the two fields: 30°, 45°, 60°; the compass needle always exactly bisects the angle between the two equal-magnitude fields.

Vector addition looks good, at least for fields of equal strength. But what about fields that have different magnitudes? Here we run into a problem, for we do not yet have a way of measuring field strength. But we might reason as follows: if one loop carrying a certain current gives a field just equal to the earth's field, then two equal loops placed right together should give twice the earth's field if magnetic fields really do add like vectors. Indeed, one loop carrying twice the current should also give twice the field strength. Let us therefore define magnetic field strength B as proportional to the current producing it, and again check on vector addition. We return to the arrangement in which the field of the loop and the field of the earth were at right angles to each other. With the original current, the compass needle stands at 45° to the direction of the earth's field; with twice this current, the angle increases to $63\frac{1}{2}°$.

This is just what we expect for the addition of perpendicular vectors having lengths in the ratio of two to one (Fig. 28 – 10). No matter how we change the current and direction of the loop, we find again and again that the direction of the total magnetic field is correctly given by the vector addition of the individual fields, if we define magnetic field strength as proportional to the current producing it. We can now represent the magnitude and direction of a magnetic field by the vector symbol $\vec{\mathbf{B}}$.



28–10 Two perpendicular vectors, with a two-to-one ratio of magnitudes, have a resultant lying at an angle of 63 1/2° from the shorter of the two vectors. If we define magnetic field strength as proportional to the current producing it, we find that magnetic fields add in this way also. Therefore, magnetic fields are vectors.

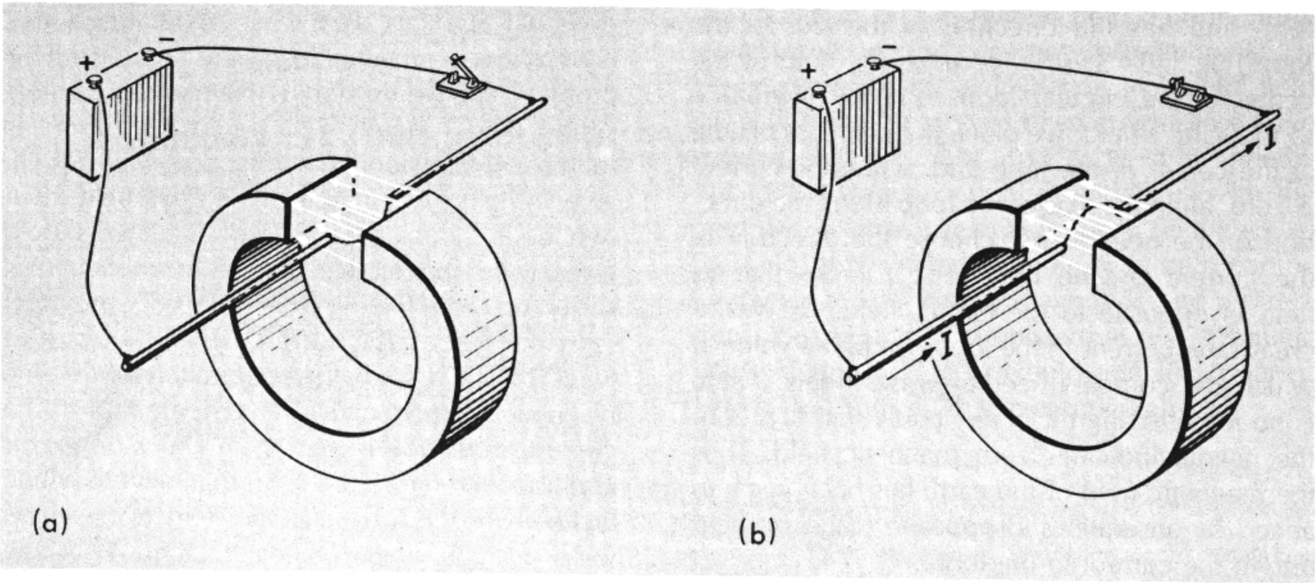
28–4 FORCES ON CURRENTS IN MAGNETIC FIELDS—A UNIT OF MAGNETIC FIELD STRENGTH

Electric charges produce electric fields which exert forces on other electric charges. Now we find that electric currents produce magnetic fields and we might guess that, correspondingly, magnetic fields exert forces on electric currents. They do, and with ordinary fields and currents the forces are quite sizable. It is these magnetic forces which spin the world's electric motors.

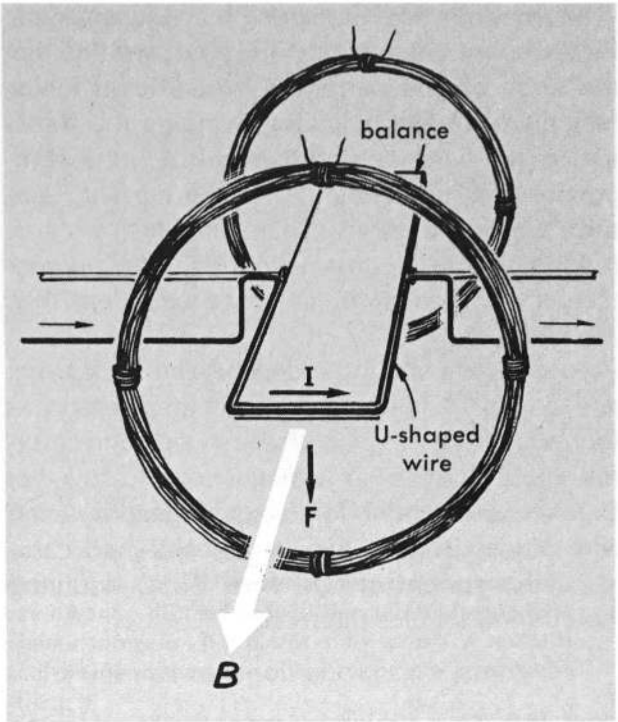
Suppose, to begin with, that we hang a flexible wire in a magnetic field (Fig. 28 – 11) The field may be either the field of a permanent magnet or the field of a current which we control. When we turn on an electric current in the flexible wire, this wire is pushed across the magnetic field.

Let us try to discover the rules describing the force experienced by a current in a magnetic field. We can find this force by measuring the force needed to hold a movable section of wire in place when a magnetic field is applied. Using experimental arrangements like that in Fig. 28 – 12, we find the following:

(i) *The force is perpendicular to both the magnetic field and the current.* Here is a simple rule to enable you to remember the actual arrangement of directions. Stretch out your *right* arm



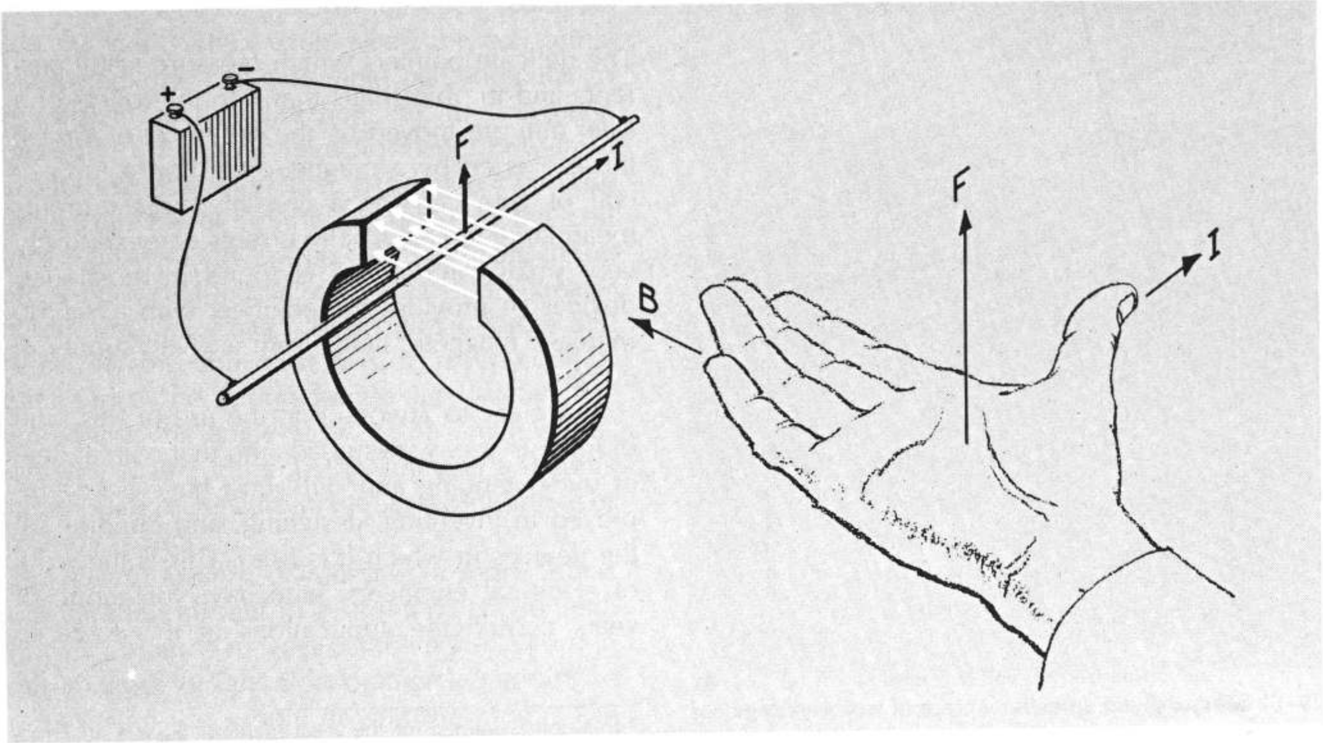
28–11 (a) A flexible wire in a magnetic field. The switch is open, no current flows, the wire hangs down. (b) When current flows, there is a magnetic force on the wire that pulls it taut.



28–12 A U-shaped wire runs on the edge of a light balance in the center of the magnetic field of two large coils. When the coils are separated a distance equal to their radii, the magnetic field at the center is almost uniform. (Such coils are called Helmholtz coils.) The straight section of wire on the end of the balance is acted on by a force $\vec{F}$ tending to tip the balance. By putting weights on the other end we measure $\vec{F}$. By turning the balance, we find how the angle between the wire and $\vec{B}$ affects $\vec{F}$.

Because the current runs in opposite directions in the two arms of the U, in a uniform field the forces on the arms cancel. Since the field is not perfectly uniform, these forces do not cancel exactly. However, the twist they apply to the balance is very small compared with the twist applied by $\vec{F}$.

with your hand open and your thumb sticking out at right angles to your fingers. Turn your right hand till the *thumb points along the current* and your *fingers point along the external magnetic field.* Then the *force* on the current is in the direction pointing *straight out from the palm of your hand;* the force is in the direction in which you would push with your hand (Fig. 28 – 13).



28–13 A rule to give the direction of the magnetic force on a current. Extend the fingers of your right hand in the direction of the magnetic field, and turn your hand so that your thumb points along the current. The magnetic force then points out from your palm — in the direction you would push with your hand.

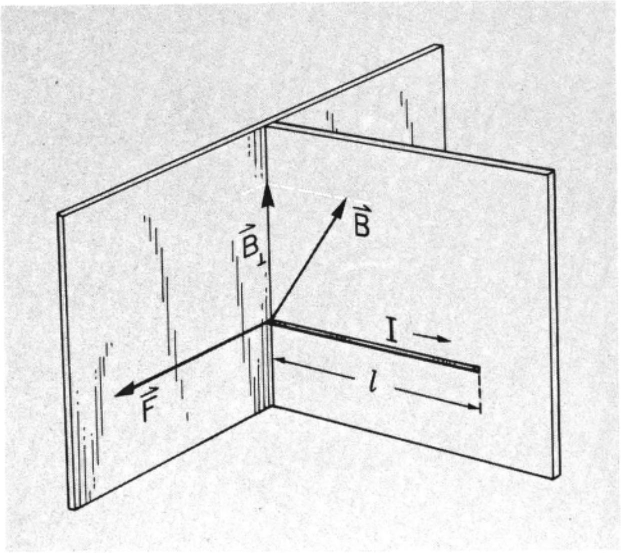
(ii) By changing the strength of the magnetic field without changing its direction, we find that the force on the current is proportional to the magnitude of the field. By changing the direction of the field we find that only $\vec{\mathbf{B}}_\perp$, the component of $\vec{\mathbf{B}}$ perpendicular to the current, contributes to the force. Therefore, *the force is proportional to* $B_\perp$, *the magnitude of the perpendicular component of the magnetic field* (Fig. 28 – 14).

To find how the force depends on the current and on the length of wire, we then reason as follows. Two wires placed side by side and carrying equal currents in a magnetic field together experience twice the force experienced by each wire alone. But the two wires combined carry twice the current of one wire. Thus, doubling the current doubles the force:

The force, F, is proportional to current, I.

Next, two short straight wires of the same length placed end to end to form one longer wire experience together double the force experienced by each length alone. Thus doubling the length of the wire doubles the force:

The force, F, is proportional to length of wire, l.



28–14 The magnetic force on a piece of wire carrying a current is proportional to $B_\perp$, the magnitude of the component of the magnetic field B perpendicular to the wire. The vector representing this force is perpendicular to the wire and to $B_\perp$; its direction is given by the right-hand rule.

To summarize our results, we now know the direction of the force and we know that the force is proportional to $B_\perp$, to I, and to l. In the form of an equation, we can write this as

$$F = (\text{const.})\ IlB_\perp$$

where the proportionality constant depends only on the choice of units.

Now we shall settle on a definite unit of magnetic field by defining it in terms of the units we have already adopted for F, I, and l. We measure F in newtons, I in amperes, and I in meters, and we choose the unit of magnetic field to make the constant equal to 1. Then

$$F = IlB_\perp$$

and the unit of B is $\dfrac{1 \text{ newton}}{(1 \text{ ampere})(1 \text{ meter})}$. A magnetic field of 1 newton/amp-meter exerts a force of 1 newton on a current of 1 ampere running perpendicularly across it for a distance of 1 meter. Another unit of magnetic field strength, frequently used in practical work, is the *gauss.** One gauss is 10^{-4} newton/amp-meter.

28–5 METERS AND MOTORS

The delicate pointers which measure small currents and the big shafts that turn the rollers of a steel mill are moved by the same sort of forces: those exerted by a magnetic field on a loop or coil of wire carrying a current. Large currents mean large forces; small currents can exert only very small forces. Iron is found in most such devices; it provides the engineer with a way to increase magnetic fields from a given supply of current.

It is clear to anyone that the use of this kind of force is very widespread, and that a great deal of understanding and skill must have been employed in inventing, designing, and building all the devices in which it is used. This is the work of electrical engineers, who have for about 75 years carried the applications of this force to constantly new uses. We are not going to try to describe the ingenious and often complex ways they have found to use the force; that is another subject. But we shall give a short account of two simple devices which depend on magnetic forces.

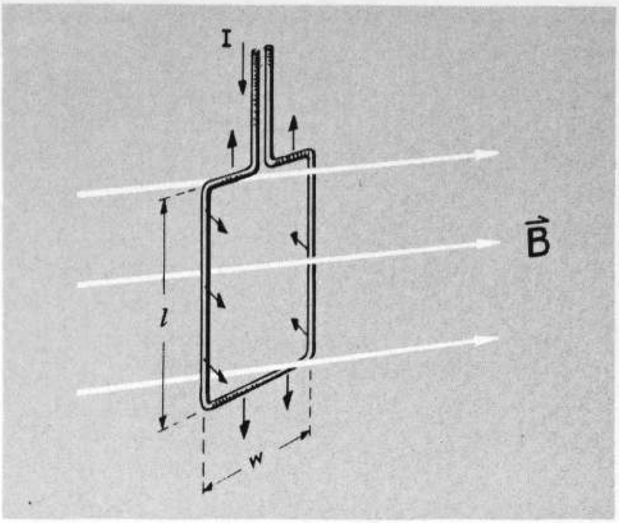
---

*This unit is named for the great mathematician Carl Friedrich Gauss, who initiated absolute measurements of the earth's field, developing a unit of magnetic field strength independent of "standard" magnets. One gauss turns out to be the field strength at the center of a circular wire loop having a radius of 2π cm and carrying a current of 10 amperes.

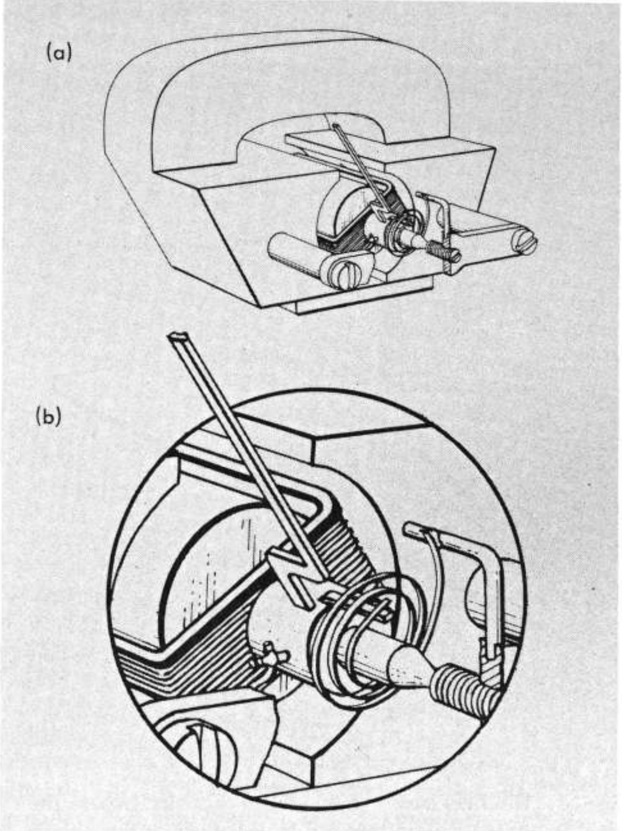
(a) *Moving-Coil Ammeters.* Figure 28 – 15 shows the forces on a loop of wire in a magnetic field. The loop coil is rectangular with sides l and w; and its plane makes an angle with the field of strength B. When a current I runs around the loop, the forces on the sides are BIl and $-BIl$. They are parallel, but they do not cancel out because they do not pull along the same line. They form a "couple" which twists the coil around the vertical axis. The forces on the top and bottom are $B_{\perp}Iw$ and $-B_{\perp}Iw$, equal and opposite. They tend to stretch the coil, but do not move it as a whole.

In an ordinary ammeter for direct currents (D.C.) the coil is carried on an axle with pivots, and its twisting is opposed by the elastic forces of a pair of hairsprings (Fig. 28 – 16). The magnetic field is produced by a permanent magnet whose field gives a radial pattern. (This is produced by shaped pole pieces and a cylindrical block of iron that is held in the open space inside the coil.) Then even when the coil turns, it still finds the magnetic field perpendicular to its sides in such a direction that the magnetic forces twist it equally hard no matter how far it has turned. On the other hand, the opposing forces due to the springs increase as the coil turns. Consequently, the coil turns until the spring forces balance the magnetic forces. The greater the current, the greater the magnetic forces and the farther the coil must twist to reach equilibrium. Since the spring forces are usually proportional to the angle of twist, this angle is proportional to the current: the ammeter's scale is linear. Such meters are made in a wide range of sizes and scales.

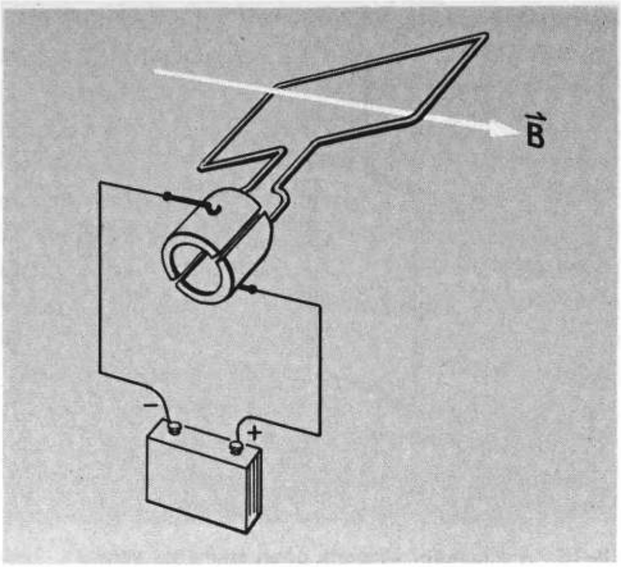
(b) *D. C. Motors.* A motor is a more robust coil, carrying current in a strong magnetic field. The D.C. motor's magnet can be either a permanent magnet or an electromagnet, carrying a current from the supply driving the motor. The coil is mounted on an axle in the magnetic field. If the current through the coil were always in the same direction, the coil would not continue to rotate; it would reach an equilibrium posi-



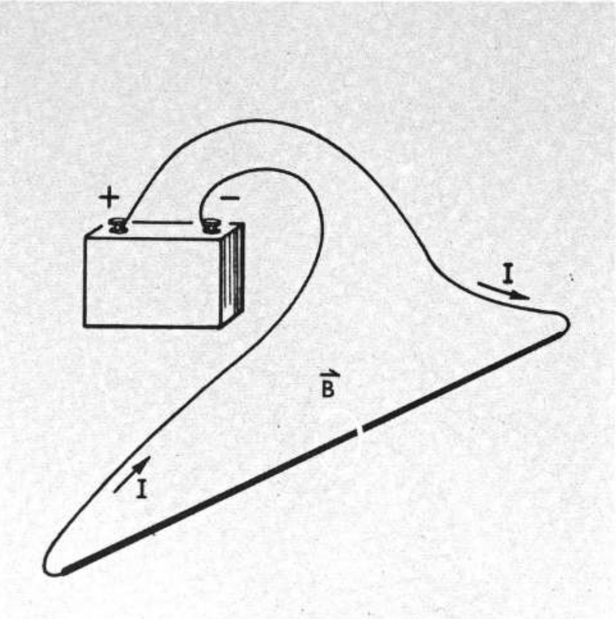
28–15 A schematic diagram of an ammeter. When current flows through the rectangular coil the magnetic force IlB and $-IlB$ tend to twist the coil. It turns until the magnetic forces are just balanced by springs. A motor is usually made so the angle of twist is proportional to the current.



28–16 (a) An ammeter. When current enters the rectangular coil of many turns through the hairspring at the front of the shaft, the coil rotates around a fixed iron cylinder. The cylinder and the specially shaped poles of the large permanent magnet give a radial magnetic field of nearly uniform magnitude. (b) Close-up of the moving coil.



28–17 A D.C. motor is similar to a moving-coil meter. The coil is bigger, and a reversing switch or commutator is provided to reverse the direction of current flow every half turn.



28–18 The field near the middle of the wire (where the field is shown) is uninfluenced by the parts of the circuit which are far away. See Fig. 28–5(b) and (c) and 28–7(a).

tion. But a motor has a device to reverse the current in the coil every half turn. The coil is pulled around for half a turn; the current is then reversed and the coil is pulled around for another half turn. One arrangement of this reversing switch, or commutator, is shown in simple form in Fig. 28 – 17. Real motors do not have a single coil but rather a whole group of coils placed so as to give steadier turning forces. Also, the coils are wound on a block of iron to increase the magnetic field.

28–6 MAGNETIC FIELD NEAR A LONG STRAIGHT WIRE

When we were studying the electric field due to a stationary charge, one of the first questions we answered was how the field falls off with distance from the charge. We have not yet asked how the magnetic field of a steady current varies with distance from the current.

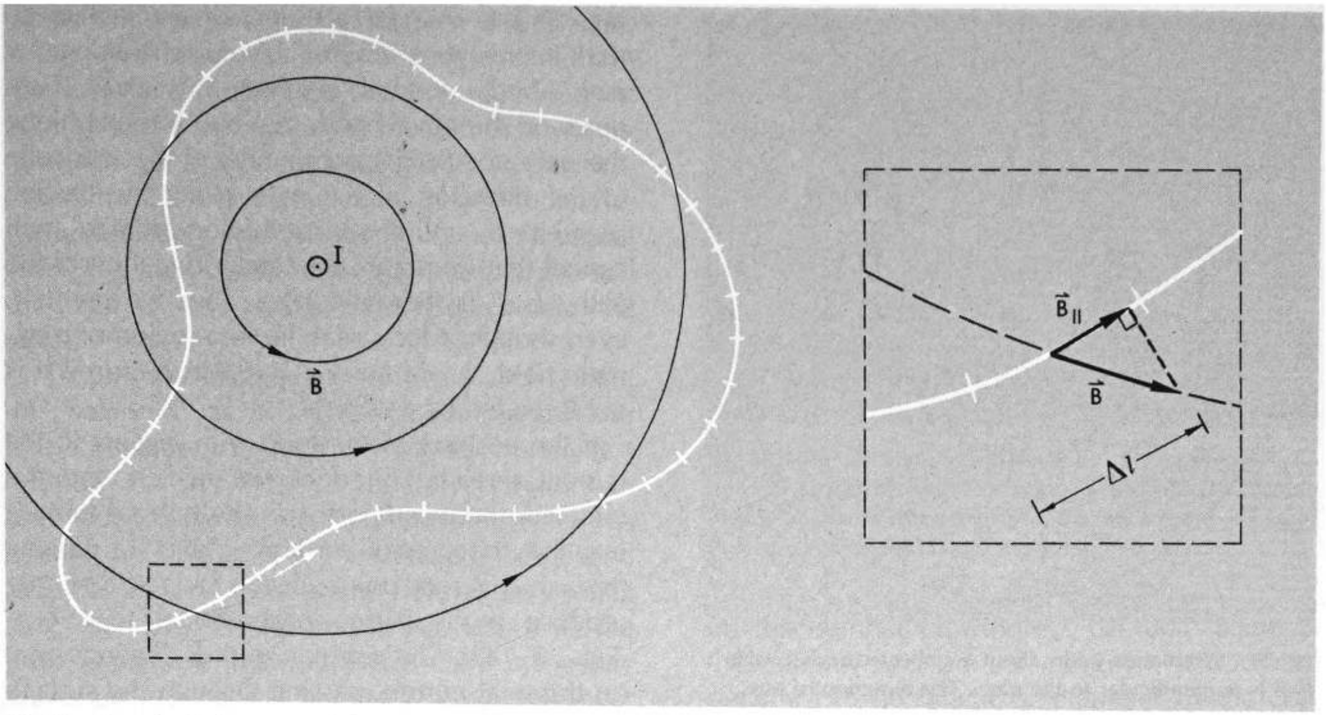
In 1820 the French physicists Biot and Savart announced the results of the first quantitative investigation of a magnetic field produced by an electric current. This was the field near a long straight wire (Fig. 28 – 18). Near the middle of such a wire, the field arises almost entirely from the current nearby. The other parts of the current do not contribute much field because the magnetic effects decrease rapidly with distance. (You can check this by moving the other parts of the circuit.) The field lines form concentric circles around the wire [Fig. 28 – 5 (b) and Fig. 28 – 7 (a)]; and the strength B of the field at a distance r from the wire was found by Biot and Savart to be inversely proportional to r. We have seen earlier in this chapter that B is directly proportional to I. Consequently,

$$B_{\text{st. wire}} = K \frac{I}{r}$$

As usual, the proportionality constant depends on the choice of units. We shall make an experimental measurement of this constant in Section 28 – 8.

28–7 CIRCULATION

An important general relation can be derived from the formula for the magnetic field of a



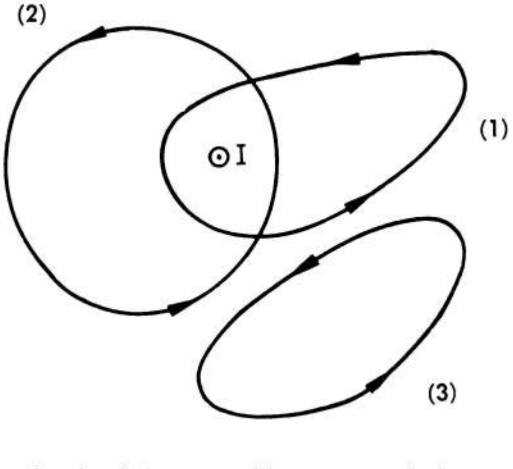
28–19 The circulation around the irregular loop is the sum around the loop of the products of each tiny segment of length Δl times $B_{\parallel}$, the component of $\vec{B}$ along the segment.

straight-line current. If we multiply both sides of the expression $B = K\,\frac{I}{r}$ by $2\pi r$ we obtain

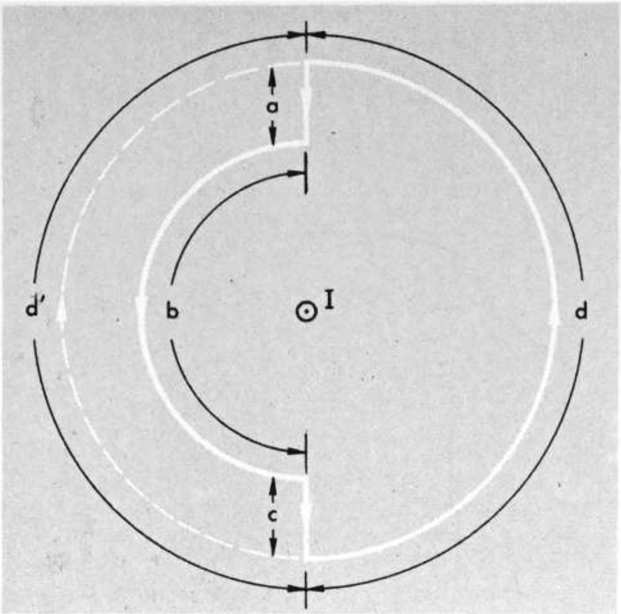
$$2\pi rB = 2\pi KI$$

The left side, $2\pi rB$, is simply the strength of the field multiplied by the length of the closed circular path to which it is tangent. And since B is inversely proportional to the radius of the circle, the product $2\pi rB$ is the same for all circles surrounding a straight-line current I. This product $2\pi rB$ is a special case of what is called the *circulation* of $\vec{\mathbf{B}}$. More generally, we may take any closed path, a smooth circle or an irregular loop, such as shown in Fig. 28 – 19, multiply the length of each small, nearly straight segment of the path by the component of $\vec{\mathbf{B}}$ along that segment and sum these products for all segments; the result is the circulation of $\vec{\mathbf{B}}$ for the path loop.

The importance of the idea of circulation is that for any loop the circulation depends only on the net current threading the loop. In Fig. 28 – 20, for example, the circulations of loop (1) and loop (2) are the same since the same current, I, threads both loops. The circulation of the adjacent loop (3), on the other hand, is zero even though it is in the magnetic field of I, since no current threads it.

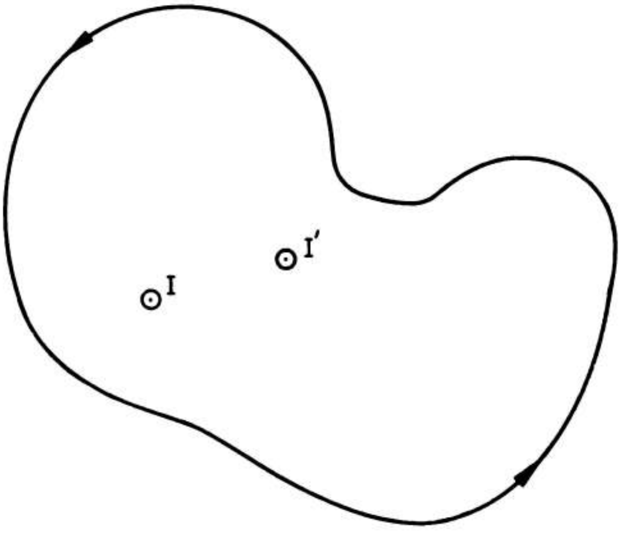


28–20 The circulation around loop (1) equals the circulation around loop (2), but the circulation around loop (3) is zero.



28-21 Circulation paths about a current-carrying wire that is perpendicular to the page. The direction of the current is out of the page, and the direction of the magnetic field is therefore counterclockwise.

To see why this is so, consider first the counterclockwise solid line of path *abcd* of Fig. 28 – 21. The contributions to the circulation of $\vec{\mathbf{B}}$ come only from segments *b* and *d*, since $\vec{\mathbf{B}}$ has no component along the radial segments *a* and *c*. The length of the arc *b* is proportional to its radius, and *B* is inversely proportional to the radius, so that the contribution to the circulation, *bB*, depends only on the angle subtended by the arc. This is true for all arcs, of any radius, so that if one goes *around* the current along the path *abcd* the total circulation is $2\pi KI$. Suppose, on the other hand, the loop is made along the path *abcd′*. The component of $\vec{\mathbf{B}}$ is opposite to the direction of the path along the dashed segment *d′* and the contribution from *d′* will cancel that from *abc*; the total circulation of the path *abcd′* is therefore zero. Thus we see that, even though a loop may lie in a region of magnetic field, it will have a zero circulation if it is not threaded by a current.



28-22 The circulation around the loop is proportional to I + I′.

If the position of the loop with respect to the current threading it does not matter, and the circulation depends only on the current threading the loop, two currents I and I' passing through different points in the loop will simply produce a circulation proportional to $I + I'$ (Fig. 28 – 22). The circulation of $\vec{\mathbf{B}}$ depends only on the total current passing through the surface bounded by the loop. It need not even be a plane surface, and the loop itself need not be planar (Fig. 28 – 23). The currents need not even be confined to long, straight narrow filaments. The general theorem relating a property of the magnetic field to currents was the culmination of the thorough investigation by Ampère (1775 – 1836) on the magnetic effects of steady currents. It is called Ampère's circuital law, or the circulation theorem for magnetic fields.

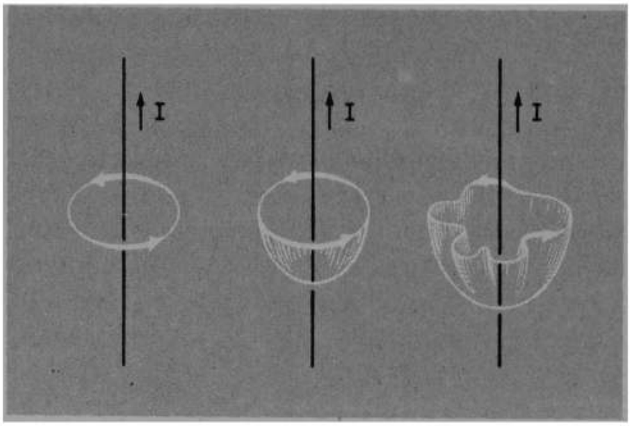
28–8 UNIFORM MAGNETIC FIELDS

To get a magnetic field that does not change with time, all we have to do is keep the current constant that generates it.* But how do we make a field that is uniform over a reasonable volume of space? From Chapter 24 you will remember that to make a uniform electric field in space we used two parallel, closely spaced metal plates carrying equal and opposite charges. With such an arrangement, the electric field was constant between the plates except near the edges.

Does the analogous prescription work here? Does a large uniform sheet of current, like a

---

*What happens when magnetic fields change with time is taken up in the next chapter.

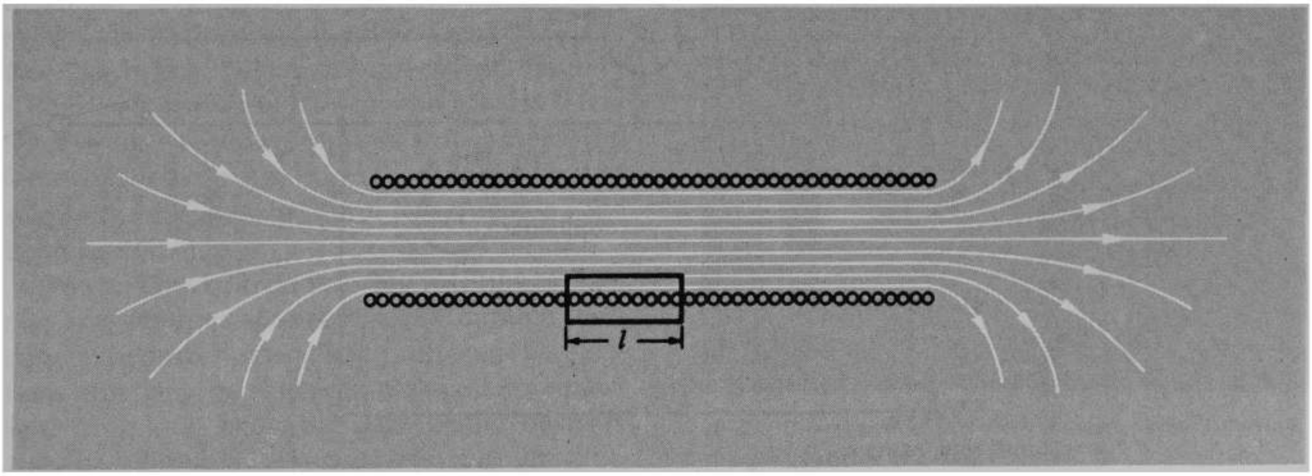


28–23 The circulation of B depends only on the total current passing through the surface bounded by the loop. This is true even when the surface is not a plane surface; indeed, even the loop itself need not lie in a plane.

large uniform sheet of charge in the electric case, give rise to a uniform magnetic field? It does. In fact, in the case of a magnetic field we can bend the plane current sheet into a cylinder and still get a uniform field inside. In practice, the current sheet is composed of one long wire wound closely and uniformly as a cylindrical coil. The field inside is along the axis of the coil [Fig. 28 – 5 (e)] and uniform except near the ends.

To find the magnetic field inside such a coil, or "solenoid" as it is often called, we can use the circulation theorem discussed in the last section. For our loop we take a rectangle with one side inside the coil and parallel to its axis where the field is $\vec{\mathbf{B}}$; the rectangle is completed outside the coil, where the field is very weak (Fig. 28 – 24). [Notice that in Fig. 28 – 5 (e) the iron filings are not noticeably lined up in any direction just outside the coil, except very near the ends.] This side therefore does not contribute anything of significance to the circulation. Neither do the two short sides, since the field lines are very nearly perpendicular to these segments, and only the component of field parallel to the path contributes to circulation. So, nearly the entire circulation comes from the length l inside the coil. We conclude, therefore, that the circulation around the rectangle is just Bl.

Suppose the rectangle threads n turns of wire. Then, from Ampère's theorem, the circulation of $\vec{\mathbf{B}}$ around the loop is $Bl = 2\pi KnI$ and the field inside is $B = 2\pi KnI/l$. That is, it depends only on the number of windings per unit length of the coil n/l and on the current I. It does not depend on the length or diameter of the coil, as long as n/l is constant and the length of the coil is large compared with the diameter. If you displace the rectangle to the right or left (but stay away from the ends) the current it threads does not change, since the coil is uniformly wound. The same holds for a displacement of the rectangle forward or backward, up or down, as long as the segment stays inside the coil. So, according to Ampère's theorem, the circulation around the rectangle does not change. Figure 28 – 5 (e) shows that the field lines inside the coil are everywhere parallel to the side l. Therefore, not only is the circulation constant, but also the field B is constant inside



28–24 To find the magnetic field inside a long solenoid, we take the circulation around a rectangle, one side of which is inside the coil and the other outside.

the coil, except near the ends. If we put $N = n/\ell$, the number of windings per unit length in the coil, the expression for the field inside the coil is $B = 2\pi KNI$.

Now, to determine the value of K in this expression all we need to do is make a measurement of the uniform field in a long solenoid having N windings per unit length and carrying a current I. We can do this with the "current balance" shown in Fig. 28 – 25 (a). The end of the balance carrying a short, straight section of wire of length ℓ' is placed in the uniform field of the solenoid [Fig. 28 – 25 (b) and (c)]. With a current I in the coil and I' in the loop, different known masses m are placed on the end of the balance outside the coil until balance is obtained. Then since $F = I'\ell' B_{\perp}$, the field B inside the coil is $mg/I'\ell'$. Equating this value for B to the expression for B we obtained using the circulation theorem, we have $mg/I'\ell' = 2\pi KNI$. This gives for the value of the constant

$$K = \frac{mg}{2\pi NII'\ell'}$$

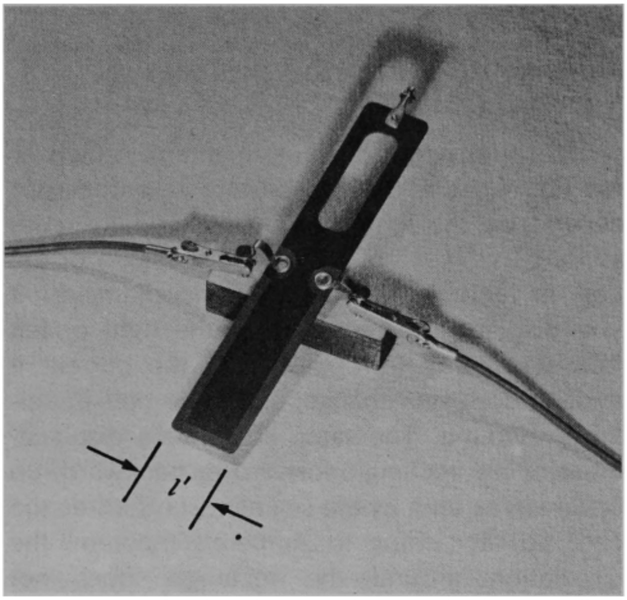
When this experiment is carefully performed, the value of K is found to be 2.00×10^{-7} nt/amp². We can now include the value of the constant in the equations for the circulation, the field about a long straight wire, and the field inside a long solenoid:

$$\text{Circulation} = (4\pi \times 10^{-7}\ \text{nt/amp}^2)\ I$$

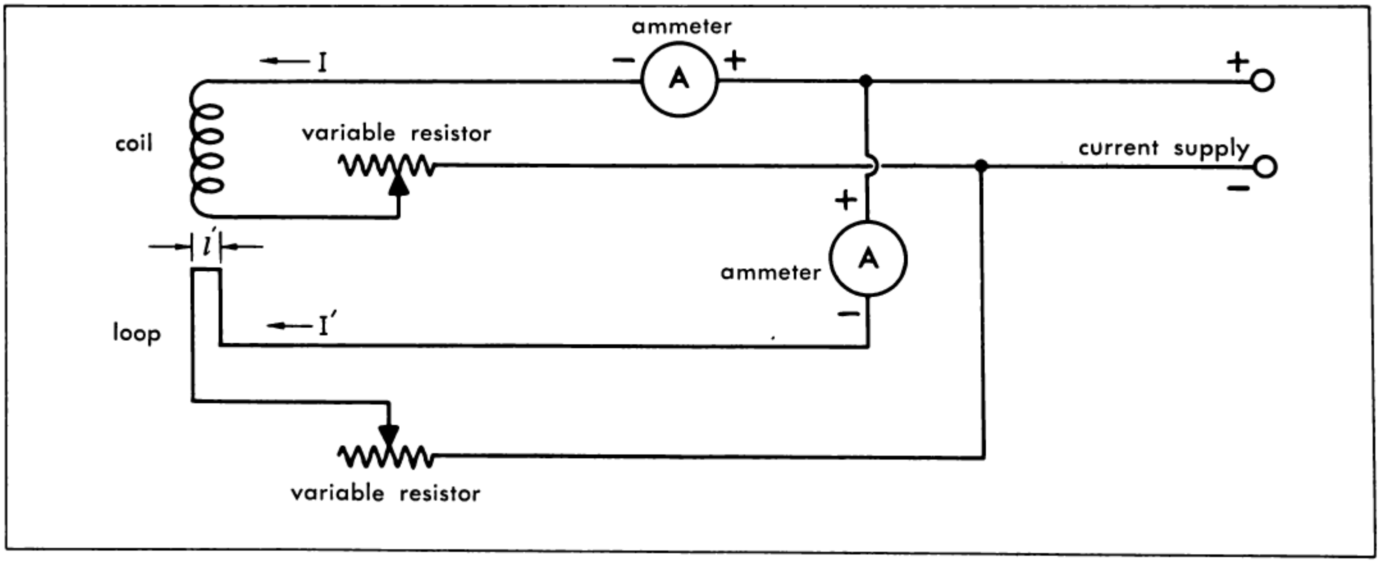
$$B_{\text{st. wire}} = (2 \times 10^{-7}\ \text{nt/amp}^2)\ \frac{I}{r}$$

$$B_{\text{solenoid}} = (4\pi \times 10^{-7}\ \text{nt/amp}^2)\ NI$$

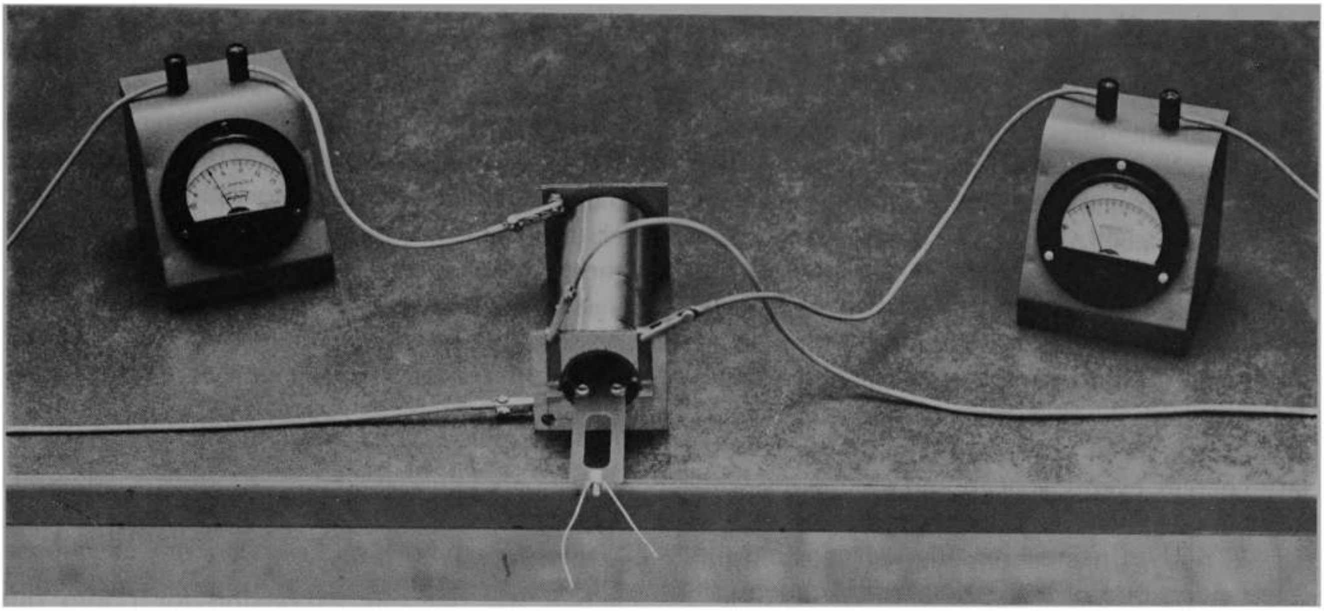
Here I must be measured in amperes and lengths in meters; then the field B comes out in $\frac{\text{newtons}}{\text{amp-m}}$ as we expect. You may wonder why the proportionality constant, K, came out to be such a simple number as $2.00 \times 10^{-7}\ \frac{\text{nt}}{\text{amp}^2}$.



28–25(a) One form of the "current balance" which can be used to determine absolute units for the magnetic field, B.



28–25(b) The far end of the current balance, carrying the short current element $I'\ell'$, is in the center of the solenoid where the magnetic field is uniform. Short pieces of string of known mass are hung on the near end to balance the magnetic force on the current element.



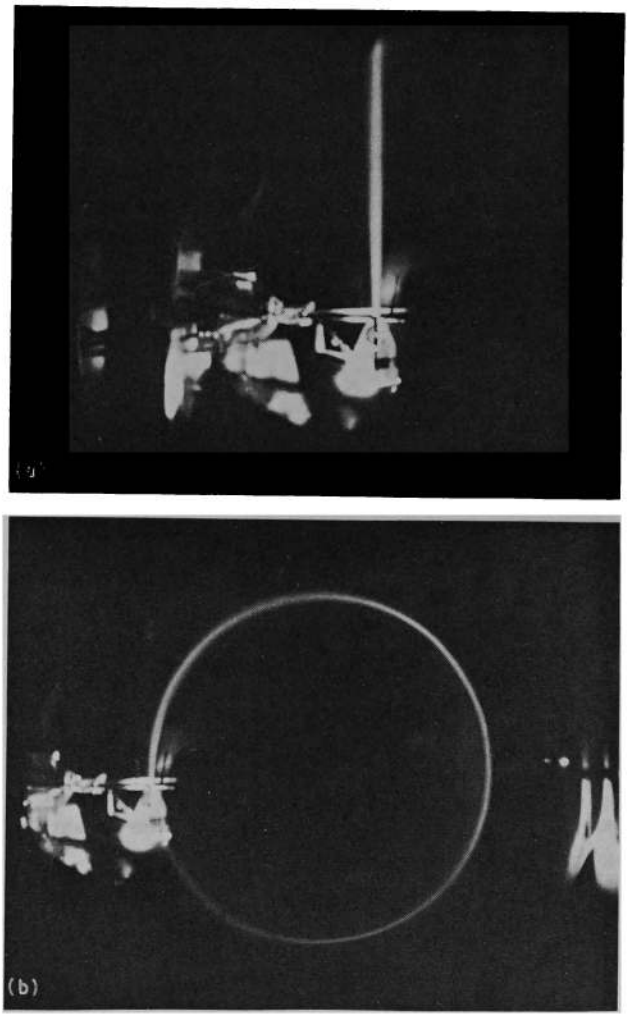
28–25(c) **The circuit connections for the experiment.**

The answer lies in the choice of units we made in Chapter 24 when we picked the ampere to be 6.25×10^{18} elementary charges per second. This made our unit for current agree with the "absolute ampere" which is usually determined by choosing K to be *exactly* 2×10^{-7} and doing an experiment similar to the one described here.

28–9 FORCES ON MOVING CHARGED PARTICLES IN A MAGNETIC FIELD

We have seen that a conductor carrying a current experiences a force when it is placed in a magnetic field. Since an electric current is a motion of electric particles, we naturally expect that the magnetic field acts directly on the individual charged particles—on the ions or electrons whose motion produces the current. The force exerted on the conductor as a whole is simply the resultant of the forces acting on the particles.

We can check up on this idea by firing a beam of electrons from an electron gun through mercury or hydrogen gas at low pressure. The electrons emerging from the slit make the low-pressure gas glow. Thus, the path of the beam is clearly visible in a darkened room (Fig. 28 – 26). Now we bring a magnet near the tube, pointed so that its magnetic field runs straight out from us across the tube. The beam is deflected at right angles to its motion and to the magnetic field.



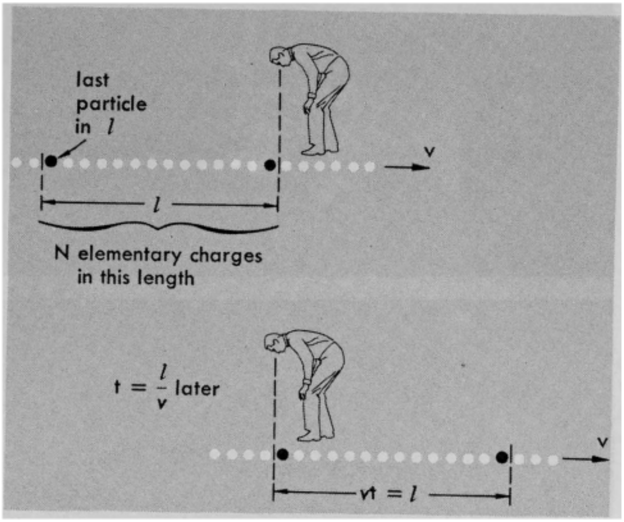
28–26 An experiment to show the deflection of moving charged particles in a magnetic field. In (a), electrons accelerated by the gun travel vertically upward, leaving a path of glowing gas. There is no magnetic field. When there is a magnetic field (b), the path of the electrons is bent into a circular arc. (We used a pair of coils rather than a bar magnet, to get a uniform field in a space into which we can look.)

To be specific, the electrons originally move vertically upward. Since electrons are negatively charged particles, this beam corresponds to a current directed vertically downward. Then, according to our rule for directions, the force acting on the moving particles must pull them to the right. In fact, as the beam goes up it is deflected to the right. If we reverse the magnet, the beam is deflected the opposite way.

We shall now compute the force acting on an individual charged particle moving with a given velocity in a known magnetic field. For this purpose we shall consider a stream of elementary charges evenly spaced and all moving with the same velocity, producing a current I (Fig. 28 – 27). Then using $F = Il B_\perp$ we can find the force F_1 on one of these elementary charges. If there are N elementary charges in the length l, then

$$F_1 = \frac{Il B_\perp}{N}$$

Thus, in order to get the force F_1 in terms of the speed v of the charges, we must evaluate the current in terms of the number of charges N in the length l. To see how I is expressed in terms of N, l, and v, we imagine an observer posted at the beginning of the length l to count the elementary charges as they go by. Because every charge is moving at speed v, all N charges in the length l stream by him in the time $t = l/v$.



28–27 An imaginary observer is watching a stream of elementary charges moving with speed/v. He counts N particles passing him when a length l of the stream goes by. This takes a time $t = l/v$.

Now, the current I is just the total charge passing per second, or N (1 elem. ch.)/t. By putting in l/v for the time, we can express this as

$$I = \frac{N\,(1\text{ elem. ch.})}{t} = \frac{N\,(1\text{ elem. ch.})}{l/v}$$
$$= \frac{Nv\,(1\text{ elem. ch.})}{l}$$

When we put this expression for I into our previous expression for F_1, we get

$$F_1 = \frac{Il B_\perp}{N} = \left[\frac{Nv\,(1\text{ elem. ch.})}{l}\right]\left[\frac{l B_\perp}{N}\right]$$
$$= vB_\perp\,(1\text{ elem. ch.})$$

As we should have expected, the expression for the force on an elementary charge is a function only of its speed and of the magnitude of the perpendicular component of the magnetic field.

If a particle moving at speed v carries n elementary charges, the force on it will be just n times as great. In other words, the force experienced by a particle carrying a charge q, equal to n (1 elem. ch.), and moving with a speed v in a magnetic field is

$$F = nF_1 = qvB_\perp$$

Let us show experimentally that this expression for the magnetic force on a moving charge is correct by shooting charged particles across a magnetic field and observing their deflection by the magnetic force. Because the magnetic force is always perpendicular to the velocity, it cannot do work on charged particles shot across the field. They, therefore, move at constant speed. Consequently, $qvB_\perp$ always has the same value. In other words, as long as the field is uniform, the magnetic force is a deflecting force of constant magnitude. The particles should go around in a circle (see Fig. 28 – 26), and we can compute the radius of this circle by equating the magnetic deflecting force with mv^2/r, the centripetal force necessary to make a particle of mass m moving at speed v go around the arc of a circle of radius r (Fig. 28 – 28). From $mv^2/r = qvB_\perp$, we find that $r = mv/qB_\perp$. Therefore, we can test our expression for the magnetic deflecting force by sending particles of known momentum and charge across a given magnetic field and measuring the radius of the circular arc on which they move.

We already have the information for protons with which we can make this test. In Section 25 – 1 we measured the mass of a proton and found it was 1.7×10^{-27} kg. We also learned that when protons are accelerated from rest through a 90-volt potential difference, they emerge with a speed of 1.3×10^5 meters per second. In this way we can produce protons of known momentum

$$mv = 1.7 \times 10^{-27}\ \text{kg} \times 1.3 \times 10^5\ \text{m/sec}$$
$$= 2.2 \times 10^{-22}\ \frac{\text{kg-m}}{\text{sec}}$$

Now we shoot test protons perpendicularly into a magnetic field. A good practical magnetic field, for example, might be 1.0×10^{-2} newtons/amp-m, and in elementary-charge units this is

$$1.6 \times 10^{-21}\ \text{newton/(elem. ch./sec)(meter)}$$

In that field the magnetic deflecting force should cause the protons to go around a circle of radius

$$r = \frac{mv}{qB_\perp} = \frac{2.2 \times 10^{-22}}{1 \times 1.6 \times 10^{-21}} = 0.14\ \text{m}$$

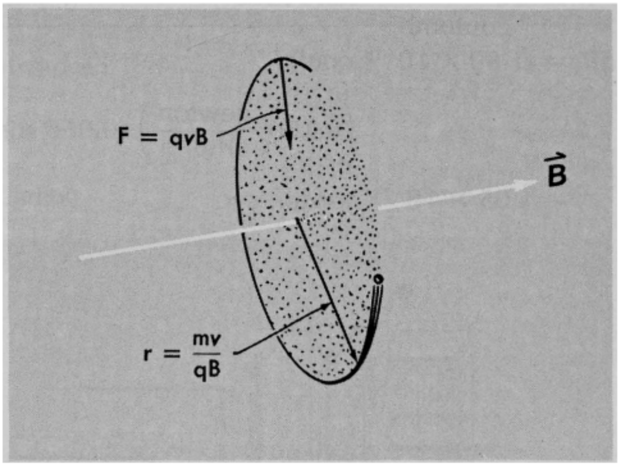
This expected radius is actually observed when protons of 90 electron-volts energy are shot into that magnetic field. Thus, experiments show that the magnetic deflecting force is indeed given by $qvB_\perp$, in agreement with our reasoning from the magnetic force on currents.

28–10 USING MAGNETIC FIELDS TO MEASURE THE MASSES OF CHARGED PARTICLES

To measure the masses of electrons and protons in Section 25 – 1, we used a known potential difference to accelerate electrons or protons to a known kinetic energy, $\frac{1}{2}mv^2$. Then we determined the speed v of these particles by doing a time-of-flight experiment.

We have another, older method of making very accurate mass measurement. Instead of measuring the speed v of the particles we can find their momentum mv by observing their deflection in a known magnetic field. If we know their kinetic energy and momentum, we can easily compute their mass. This is the method by which our most precise measurements of the masses of ions have been made in the past. It is particularly valuable for electrons because their high speeds make time-of-flight measurements difficult.

The measurement of momentum by deflecting a beam of particles in a magnetic field is basically the same measurement that we made at the end of the last section. There we used the deflection of protons as an experimental check on the expression $F = qvB_\perp$ for the magnetic force. Here we shall use the known magnetic force to find unknown masses.



28–28 When a charged particle moves across a magnetic field, it experiences a magnetic force F = qvB at right angles to its path. If the field is uniform, the particle moves in a circle of radius r = mv/qB.

To use this method to measure ionic masses, we accelerate the ions of unknown mass through a known electrical potential difference, V (Fig. 28 – 29). This tells us their kinetic energy: $\frac{1}{2}mv^2 = qV$. Then we allow the ions to go perpendicularly into the known magnetic field and measure the radius of the circular arc on which they move. The required centripetal force mv^2/r is supplied by the magnetic deflecting force $qvB_\perp = qvB$. Thus, $mv^2/r = qvB$, and so we get $mv = qBr$. Consequently, by measuring the radius of the circular arc, we can evaluate the momentum of the particle.

Let us take a specific example. We accelerate singly charged sodium ions through an electric potential difference of 90 volts. Any singly charged ion accelerated through 90 volts acquires 1.44×10^{-17} joule of kinetic energy. With this kinetic energy the sodium ions enter a uniform magnetic field of strength 0.100 newton/(amp-meter) at right angles to their direction of motion. Now we measure the radius of the arc through which they travel in the magnetic field, and we find that it is 0.066 meter. Since the ions are singly charged, each with 1.60×10^{-19} coulomb,

$$qBr = (1.60 \times 10^{-19}\ \text{coul})\left(0.100\ \frac{\text{newton}}{\text{amp-m}}\right)(0.066\ \text{m})$$
$$= 1.05 \times 10^{-21}\ \text{newton-sec}$$

This is the momentum of the ions.

Now from the kinetic energy $\frac{1}{2}mv^2$, and the momentum mv, we can get both m and v. The speed is easily determined by

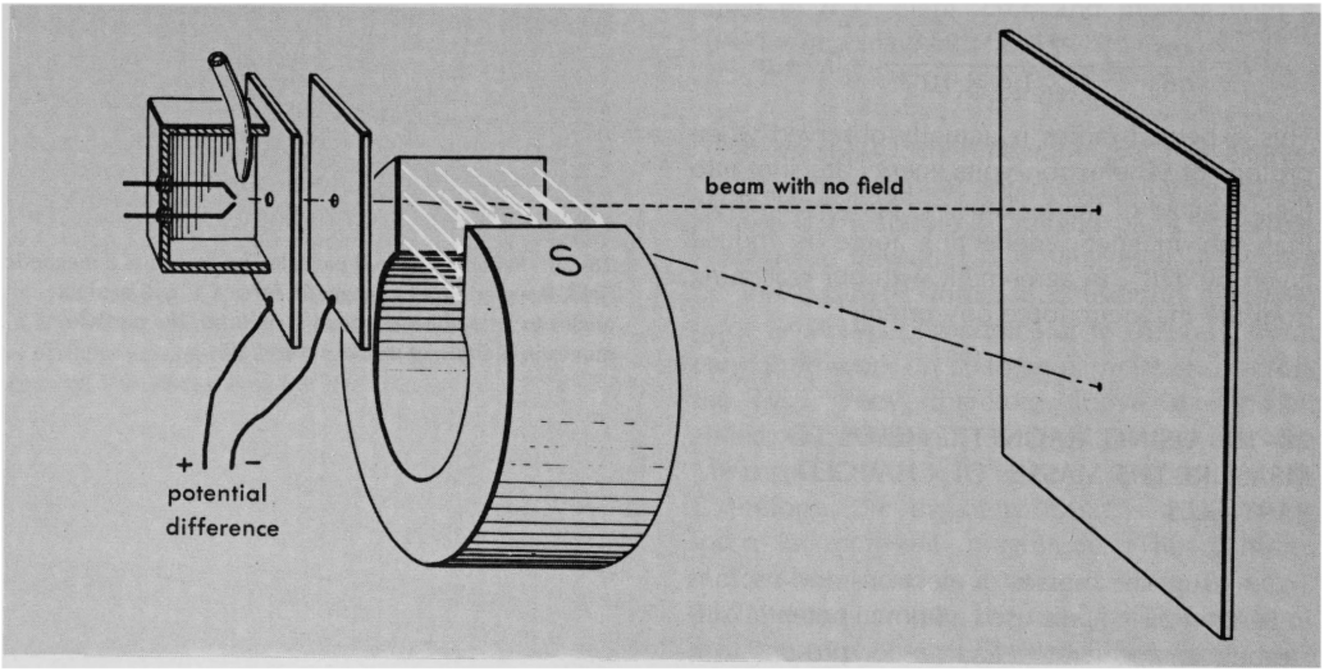
$$v = \frac{2(\frac{1}{2}\ mv^2)}{mv} = \frac{2(1.44 \times 10^{-17}\ \text{joule})}{1.05 \times 10^{-21}\ \text{newton-second}}$$
$$= 2.74 \times 10^4\ \text{meters/second}$$

Finally, dividing this speed into the momentum, we get the mass of the ions:

$$m = \frac{mv}{v} = \frac{1.05 \times 10^{-21}}{2.74 \times 10^4} = 3.84 \times 10^{-26}\ \text{kg}$$

This is the mass of sodium ion – good to within a percent. The mass of a sodium atom, which is made by adding one electron to the sodium ion, must be almost identical. So our value can be compared with the values obtained for the mass of sodium in other experiments. The accepted value for the mass of a sodium atom is 3.82×10^{-26} kg.

The apparatus that we have diagrammed in Fig. 28 – 29 is known as a mass spectrograph, and the procedure of measuring masses with such apparatus is called mass spectroscopy. Our mass spectrograph is very crude. One that is made for high-precision measurements usually contains auxiliary devices to "focus" the particles on a small spot and to make sure that only particles in a very narrow range of velocities can enter the magnetic field. With a good



28–29 A simple mass spectrograph. Ions are accelerated to a known kinetic energy by the electric field between the plates. In the magnetic field their path is bent into a circular arc. The ions strike a screen or photographic plate, and from the deflection we can calculate the radius of the arc.

spectrograph masses of ions can be determined to many significant figures.

However, such measurements show that the masses of the elements are not exact integral multiples of the mass of the hydrogen atom. A precision mass spectrograph shows that many elements are composed of two or more kinds of atoms, differing in mass but having almost identical chemical properties; for example, 75 percent of the chlorine atoms found in nature have a mass of about 35 times the hydrogen atomic mass, while 25 percent have a mass of about 37 times the hydrogen atomic mass. Atoms of the same element which differ in mass we call different *isotopes* of the element. The two chlorine isotopes are called Cl^{35} and Cl^{37}. The superscript (called the *mass number*) labelling an isotope is an indication of the internal structure of the atoms of the isotope.

An isotope of carbon, C^{12}, has been chosen as the standard for the currently used scale of *atomic mass units.** Its atomic mass is defined to be precisely 12 atomic mass units; on this scale hydrogen has an atomic mass of 1.00797 amu and oxygen has an atomic mass of 15.9994 amu, for example (Appendix 3). You will notice that the atomic mass of carbon is listed as 12.01 amu; this value, like the others in the table, is the average atomic mass of the element, when all isotopes are considered in their natural abundances. The C^{13} isotope, which has an abundance of 1.1 percent, raises the average atomic mass of carbon, as it occurs in nature, to a value slightly greater than an even 12 amu.

One advantage of the C^{12} scale over a hydrogen-based scale is that the mass of any isotope in amu is then very close to the mass number of that isotope. For example, the uranium isotope U^{238} has an atomic mass of 238.03 amu on the C^{12} scale, although it has only 236.20 times the mass of a hydrogen atom.

Mass spectrographs can be used also to measure the mass of molecules by deflecting a beam of ionized molecules in the apparatus just as we deflect a beam of ionized atoms.

Ever since J. J. Thomson introduced mass spectroscopy about fifty years ago, it has been a powerful tool of atomic research. In particular, because mass spectrographs sort the ions by their individual masses, the masses of the isotopes of a single element were largely determined by mass spectrographic measurement. Although some evidence for the existence of isotopes among the heavy radioactive elements was pointed out in the first few years of this century, it was Thomson's measurement showing the existence of the isotopes of neon that clinched the idea and started the accumulation of systematic data on isotopes. By now we have accumulated so much information that a huge table would be needed to exhibit it. Only a small sample of the known masses of isotopes is shown in Table 1.

TABLE 1

A few of the isotopes of some common elements. One atomic mass unit is $\frac{1}{12}$ of the atomic mass of the most common carbon isotope.

ELEMENT	MASSES OF THE ISOTOPES (ATOMIC MASS UNITS)	RELATIVE ABUNDANCE (PERCENT)
Hydrogen	1.008	99.98
	2.014	0.02
Lithium	6.015	7.42
	7.016	92.58
Carbon	12.000 (standard)	98.89
	13.003	1.11
Nitrogen	14.003	99.63
	15.000	0.37
Oxygen	15.995	99.76
	16.999	0.04
	17.999	0.20
Neon	19.992	90.92
	20.994	0.26
	21.991	8.82
Sodium	22.999	100
Magnesium	23.985	78.70
	24.986	10.13
	25.983	11.17
Chlorine	34.969	75.53
	36.966	24.47
Iron	53.940	5.82
	55.935	91.66
	56.935	2.19
	57.933	0.33
Silver	106.905	51.82
	108.905	48.18
Lead	203.973	1.5
	205.975	23.6
	206.976	22.6
	207.977	52.3

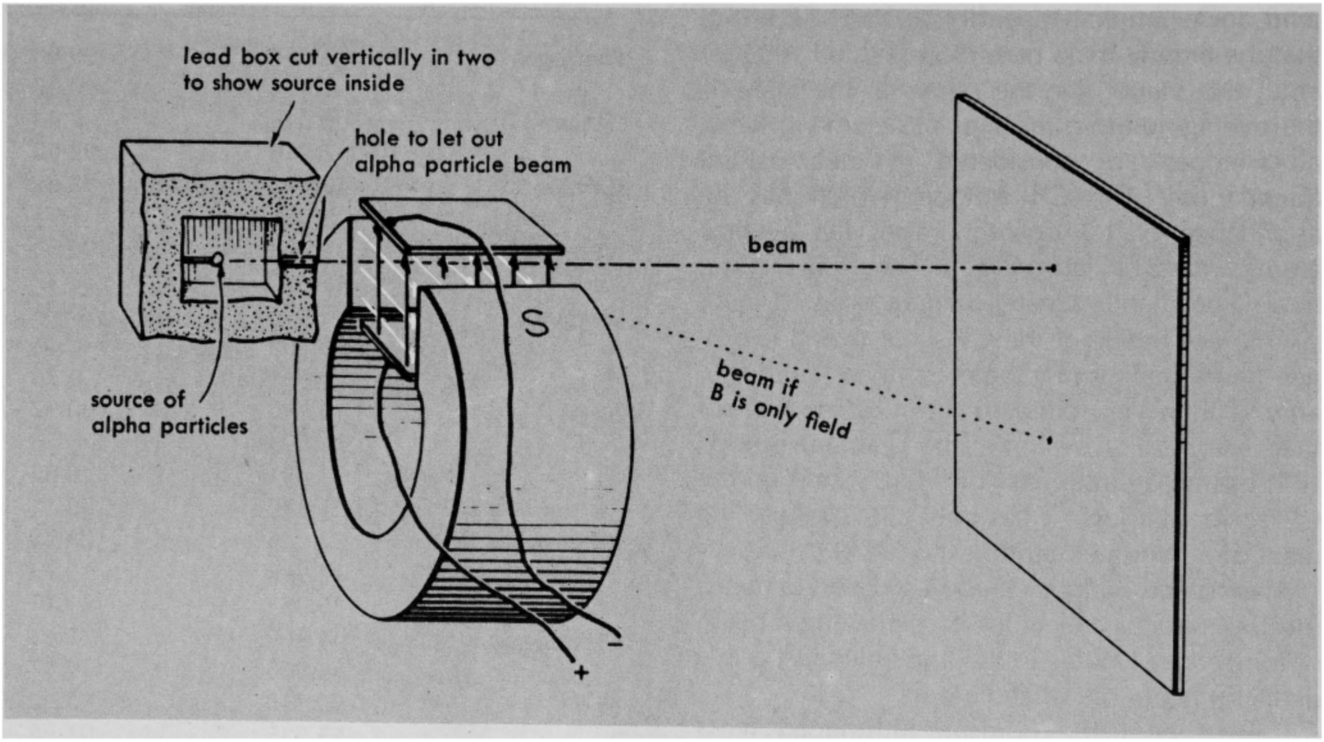
---

*Incidentally, most of the tables of atomic masses printed before 1961 were based on oxygen as an even 16 amu. Due to a number of serious difficulties, the oxygen scale was officially abandoned in 1961 and replaced by the carbon-12 scale.

28–11 WHAT ALPHA PARTICLES ARE

In Chapter 26 we saw some of the evidence about the structure of atoms accumulated by Rutherford and his co-workers in the first few decades of this century. Rutherford's chief experiments were done with alpha particles. We also know how Rutherford and Geiger determined that alpha particles carry two elementary charges (see Section 24 – 9). Now we shall describe two experiments equivalent to those done by Rutherford and Robinson to determine the mass of the alpha particle. In one of these experiments we determine the momentum of an ion by measuring its deflection in a magnetic field. Because we have just described this kind of experiment in the last section, we only mention the result. When we do this experiment using polonium as the source of alpha particles, we find that their momentum is $mv = 1.06 \times 10^{-19}$ kg-meter/sec.

Now, to obtain the mass of the alpha particle, we must measure its speed. We could make this measurement by stopping the alpha particles with a retarding electric potential difference, thus determining their kinetic energy. This would require a potential difference of a million volts or so. But we can get the speed more directly by applying an electric field across the motion of the alpha particles at the same time that we apply a magnetic one. We put on this electric field so that the electric force on the alpha particles is just opposite to the magnetic force. Then the net force on the moving particles is zero, and they move straight ahead, following the same path that they take in the absence of both fields (Fig. 28 – 30). The electric force on the particles is qE, and the magnetic force qBv; and when these forces balance, $qBv = qE$. The speed of the particles is therefore given by $v = \frac{E}{B}$, and since we know both the magnetic and the electric field, we know the answer. For polonium alpha particles the meas-



28–30 Measuring the speed of alpha particles. With a pair of metal plates placed horizontally in the tube we can apply an electric force equal and opposite to the magnetic force. Then the particles are not deflected, and their speed v = E/B. This figure is schematic. In a real experiment, we must change the dimensions to obtain known uniform electric and magnetic fields. Also, the apparatus is in a vacuum system to avoid scattering and energy loss by alpha particles hitting air molecules.

ured speed is $v = 1.6 \times 10^7$ meters per second. Consequently, the mass is

$$m = \frac{mv}{v} = \frac{1.06 \times 10^{-19}}{1.6 \times 10^7} = 6.6 \times 10^{-27}\ \text{kg}$$

Within the experimental accuracy, this mass is almost exactly equal to the mass of a helium atom. It is thus natural to conclude that alpha particles are doubly charged helium ions. We can test this conclusion directly by placing a strong source of alpha particles in an evacuated glass tube. Such an experiment was done by Rutherford and Royds in 1908. After waiting for a sufficiently long time, they found a detectable amount of helium.

Looking back at the contents of this and the previous chapters, you will realize that we have dealt so far with electric fields produced by electric charges at rest, and with magnetic fields produced by steady currents. Such fields are constant in time. In the next chapter we shall again turn to experiment to learn what happens when electric and magnetic fields vary with time.

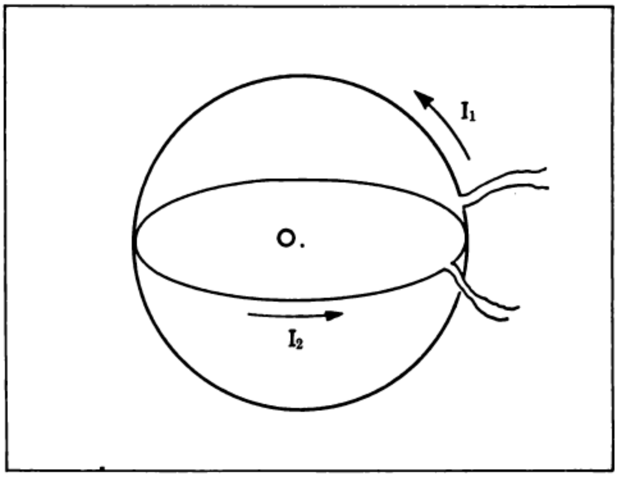
QUESTIONS

1.* If a long straight wire carrying a current were placed flat on a paper and iron filings were sprinkled on the paper, what would you expect the iron filings to do?

2.* Do magnetic field lines have a beginning and ending?

3.* What is the direction of the magnetic field produced by positive charges moving directly away from you? Negative charges moving toward you?

4.* Two identical coils carry equal currents, have a common center, and are at right angles to each other. How does the magnitude of the field they produce compare with the magnitude of the field from just one of them?



28–31 For Question 7.

5.* How does the field produced by a current-carrying wire depend on the magnitude of the current?

6. A horizontal uniform magnetic field $\vec{B}$ is perpendicular to the horizontal component $\vec{B}_e$ of the earth's magnetic field and points east.
   (a) The ratio $B/B_e = \sqrt{3}$; in which direction will a compass needle point?
   (*Note:* the compass needle rotates in a horizontal plane.)
   (b) If the compass needle points northeast, what is the magnitude of $\vec{B}$ now?

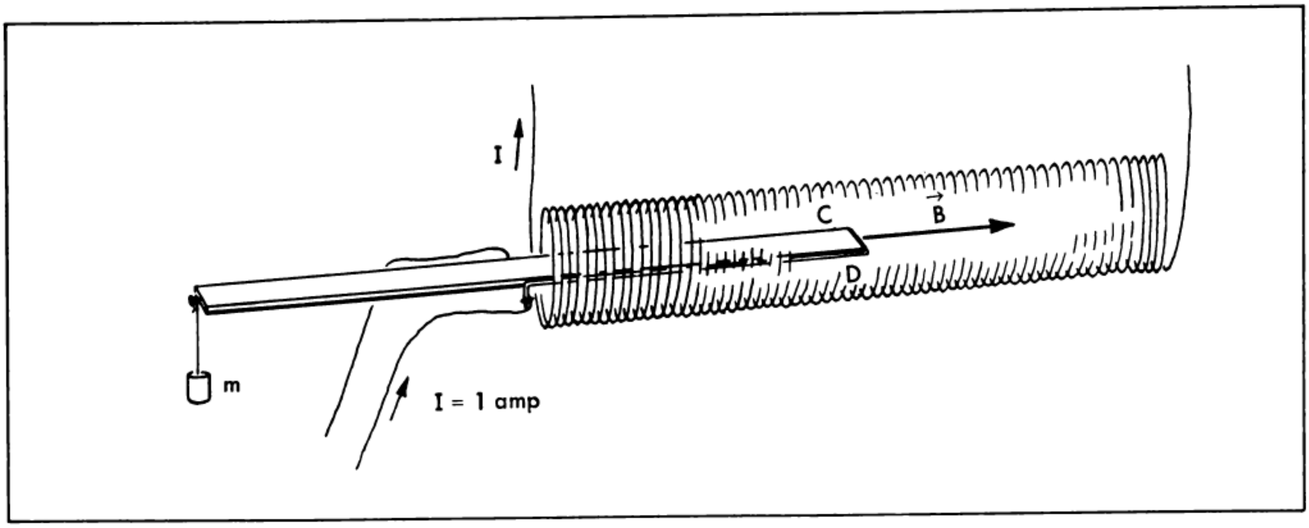
7. Two circular wire loops of equal radius are set up at right angles to each other with a common center (Fig. 28 – 31). If $I_1 = 3.0$ amperes and $I_2 = 4.0$ amperes, what is the direction of the magnetic field at the center, *0?*

8. Suppose two loops are fixed at right angles and you can run a current in either direction through either loop. Neglecting the earth's field, how many different magnetic fields could you produce if a current of 2 amps flows whenever a loop is connected?

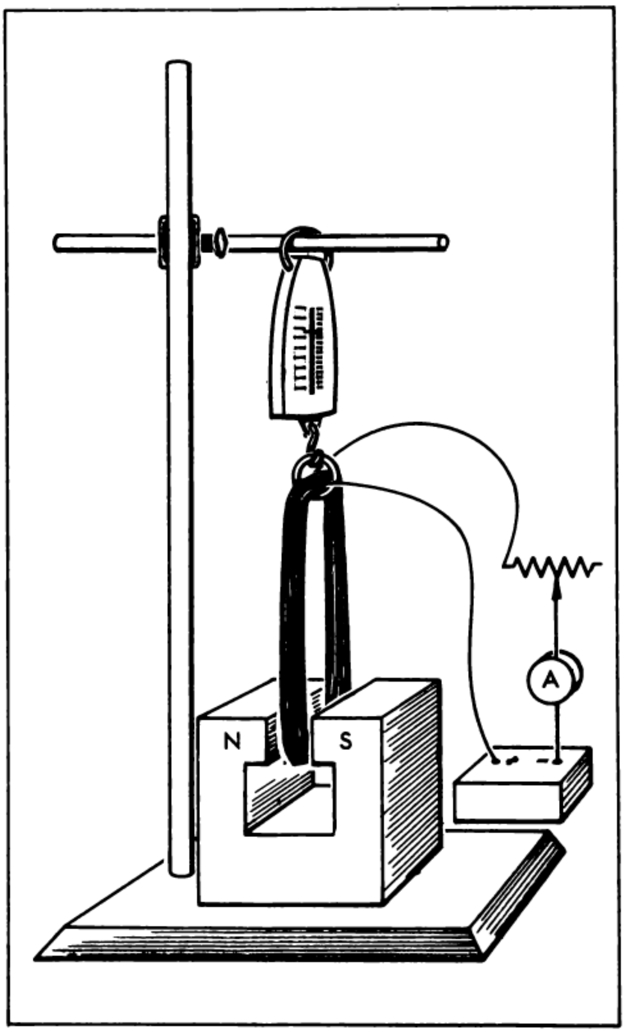
9.* In what direction can an electric charge move through a magnetic field without having a force exerted on it?

10.* Suppose a wire carries a current from left to right through a magnetic field whose direction is toward you. In what direction is the magnetic force on the wire?

11.* What is the force on a wire 0.20 meter long, carrying a current of 3.0 amps, which is placed within a uniform magnetic field of 5.0 newtons/amp-m at an angle of 45° to the field?



28–32 For Question 12.



28–33 For Question 15.

12. The magnetic field inside a long solenoid is uniform and parallel to the axis of the solenoid.

(a) What is the force on a current-carrying wire inside the solenoid and parallel to the axis?

(b) What is the force on the wire shown in Fig. 28 – 32 which carries a current of 1.0 ampere, if the field B in the solenoid is 1.0×10^{-2} newton/amp-m and the length CD is 2 cm?

(c) How much mass m should be placed on the other end of the balance to keep it from tilting?

(d) Describe how you could use this setup as an ammeter.

13. (a) Show that a square loop of wire carrying a current will always tend to align itself in a magnetic field so that the plane of the loop is perpendicular to the field.

(b) What do the magnetic forces tend to do to the loop in this position? (Hint: draw a side view.)

14. A 50-turn coil is placed in a magnetic field as in Fig. 28 – 15. Both l and w are 1 cm. $B = 0.10$ newtons/ampere-meter and is uniform over the entire region occupied by the coil.

(a) Find the magnitude of the force on each side of the coil when the current is 5.0×10^{-3} amperes and the plane of the coil is at 45° to the direction of the magnetic field. What is the direction of the force on each side of the coil when the direction of the current is down on the near side of the coil?

(b) If the coil is free to rotate about a vertical axis, in what position should the coil be placed so that a current in the coil will not cause the coil to turn?

15. In the moving-coil ammeter of Fig. 28 – 16, the magnetic force on the coil is balanced by a torsion spring. An ammeter can also be constructed as shown in Fig. 28 – 33, using a small spring balance to measure the magnetic force on the coil. By adjustment of the position of the support, the conductors can be kept in the same position in the magnetic field for different currents.

One such coil, 4.0 cm wide by 25 cm high, was constructed of 100 turns and weighed 2.4 newtons. The spring balance had a scale of 0 to 5 newtons.

(a) When supported in a magnetic field, as shown, the spring balance changed from 2.4 to 2.7 newtons as the current was increased from 0 to 0.5 ampere. What is the maximum current that could be measured on this spring balance?

(b) What is the magnetic field in the air gap of the magnet?

16.* How does the force that opposes the magnetic force in a moving-coil ammeter depend on the angle of rotation of the pointer on the ammeter's scale?

17.* How could the direction of rotation of the coil in Fig. 28 – 17 be reversed?

18.* What happens to the radius of the circle in Fig. 28 – 26 (b) if the accelerating potential inside the electron gun is raised?

19.* Upon what does the magnetic force on a moving electric charge depend?

20. How do the momenta of a proton and an electron compare if both are moving perpendicular to the same magnetic field in circles of the same radius?

21. Would it be easier for a mass spectrograph to separate the isotopes of light elements or heavy elements?

22. A beam of protons of kinetic energy E_K is injected into a region of uniform magnetic field $\vec{\mathbf{B}}$ at right angles to the direction of the beam.

(a) Find expressions for the radius and the period of the resulting circular motion in terms of the magnetic field strength.

(b) How would your answers to (a) be affected if the incident energy were four times as great?

(c) Protons of energy 4.60×10^5 electron volts are injected into a proton synchrotron in which the magnetic field strength is 2.10×10^{-2} newton/amp-m. Calculate the period and the radius of the circular motion.

23.* What is the kinetic energy of alpha particles accelerated by a potential difference of 180 volts?

24. Charged particles are accelerated by a uniform electric field and then deflected by a magnetic field. What quantities must be measured to determine the mass of the particles from this experiment?

25. When helium gas is fed into the apparatus shown in Fig. 28 – 29, mounted in a chamber from which the gas can be pumped to keep the pressure low, He^+ and He^{++} ions are produced. These two ions form two spots on the screen below the position of the beam with no field. The face of the magnet is 3.0 cm by 1.5 cm and the center of the face is 10 cm from the screen. The strength of the magnetic field is 1.5×10^{-2} newton/amp-m.

(a) Compute the radius of the circular part of the path of each ion while it is in the magnetic field.

(b) Make a scale drawing and determine the position of each spot on the screen.

(c) What would be the path of a He^{++} ion which gains one electron while in the electric field? Describe it qualitatively.

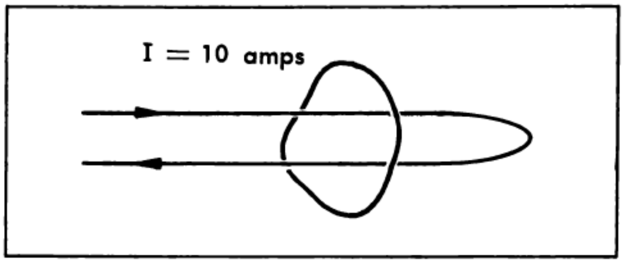
(d) What would be the path of a He^{++} ion which gains one electron while in the magnetic field? Describe it qualitatively.

26. A nearly uniform electric field is set up parallel to the surface of a glass-topped table by charging two large parallel, vertical metal plates. The plates are 0.50 m apart and the potential difference between them is 20 kv. A small Dry Ice puck with a total mass of 60 grams carries a copper-coated Ping-Pong ball which has a charge of 2.2×10^{-8} coulomb.

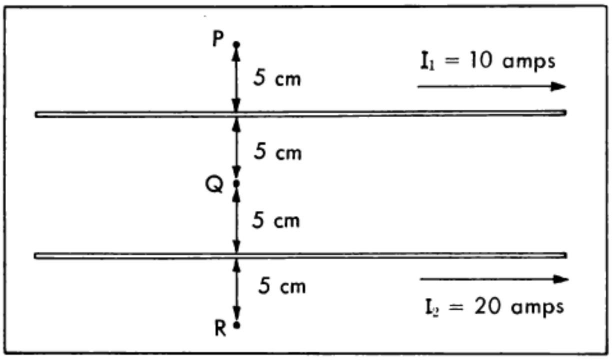
(a) What is the electric force on the ball when it is placed between the plates?

(b) What would be the shape of its path if it were given a brief push perpendicular to the field lines?

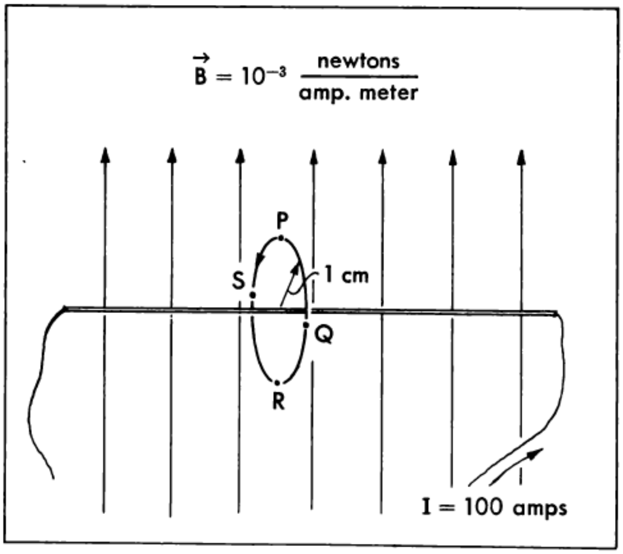
(c) Even in a restricted region, the largest magnetic field that can be obtained fairly easily is 0.50 newton/amp-m. If you projected the puck through such a field perpendicular to the field at a speed of one meter/second, what would be the magnetic force on the puck?



28–34 For Question 31.



28–35 For Question 33.



28–36 For Question 34.

27. A beam of singly charged ions moves into a region of space where there is a uniform electric field $E = 1.0 \times 10^3$ newton per coulomb, and a uniform magnetic field $B = 2.0 \times 10^{-2}$ newton/amp-m. The electric and magnetic fields are at right angles to each other and both are perpendicular to the beam so that the electric and magnetic forces on an ion oppose each other. What is the speed of those ions which move *undeflected* through these crossed electric and magnetic fields?

28. The undeflected ions of Question 27 are passed through a slit and move into a region where there is a uniform magnetic field $B = 0.09$ newton/amp-m at right angles to their motion. If the ions are a mixture of the neon ions of mass 20 and 22 atomic mass units, how far apart will these ions land on a photographic plate after they have moved through a semicircle?

29. A long, straight wire at rest in the laboratory frame of reference carries an electric current I, which consists of electrons moving with speed v from left to right through the wire. An observer also at rest in the laboratory frame notes a magnetic field whose magnitude is given by $B = K\frac{I}{r}$.

    (a) If the observer walks parallel to the wire from left to right with speed v, will he observe the same magnetic field as he did when he was at rest?

    (b) If the observer is at rest but the whole wire moves from left to right with speed v relative to the laboratory (still carrying electrons moving from left to right with speed v with respect to the wire), will he observe the same magnetic field as he did when both he and the wire were at rest?

30.* A wire carrying a current I threads a square loop 2 cm on a side. What is the circulation around the square? (Express your answer in terms of K.)

31.* What is the circulation around the loop shown in Fig. 28 – 34?

32.* What is the field in the center of a long solenoid having 10 turns/cm and carrying a current of 5 amp?

33. Two very long straight parallel wires, one carrying 10 amperes and the other 20 amperes, are separated by a distance of 10 cm (Fig. 28 – 35). Find the magnitude and direction of $\vec{\mathbf{B}}$ at the points P, Q, and R shown in the figure.

34. A long straight wire carrying a current $I = 100$ amperes is perpendicular to a uniform magnetic field of strength $B = 1.0 \times 10^{-3}$ newton/amp-m (Fig. 28 – 36). What is the strength and direction of the resultant $\vec{\mathbf{B}}$ field at the points P, Q, R, S on a circle of radius 1 cm around the wire?

35. Suppose we have a beam of electrons moving at a speed $v = 3.00 \times 10^6$ m/sec and carrying a current of 1.00 microampere.

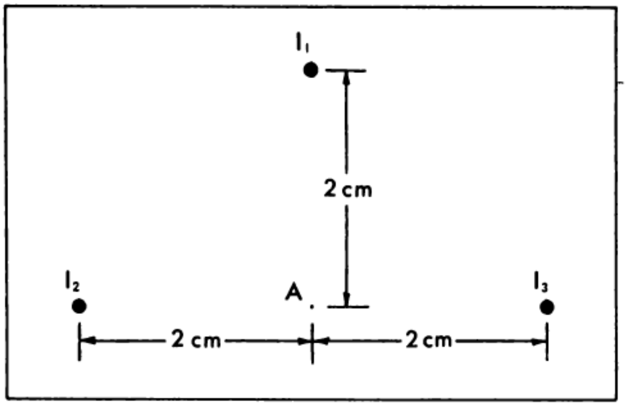
(a) How many electrons per second pass a given point?

(b) How many electrons are in 1.00 meter of the beam?

(c) What magnetic field does the beam produce at 1.00 m distance?

(d) What is the total force on all of the electrons in 1.00 m of the beam if it passes through a field of 0.10 newton/amp-m?

(e) What is the force on a single electron if you assume that each electron experiences the same force?



28–37 For Question 36.

36. Point A is 2 cm from each of 3 currents I_1, I_2, and I_3, as shown in Fig. 28 – 37. All three currents are carried by long wires perpendicular to the paper.

(a) If $I_1 = I_2 = I_3 = 10$ amperes, and all three currents are directed out of the paper, what is the magnitude and direction of the magnetic field at A due to the currents in the wires?

(b) If $I_1 = I_2 = I_3 = 10$ amperes, but I_2 is into the paper while I_1 and I_3 are out, what is the magnitude and direction of the magnetic field at A due to the currents in the wires?

29 Electromagnetic Induction and Electromagnetic Waves

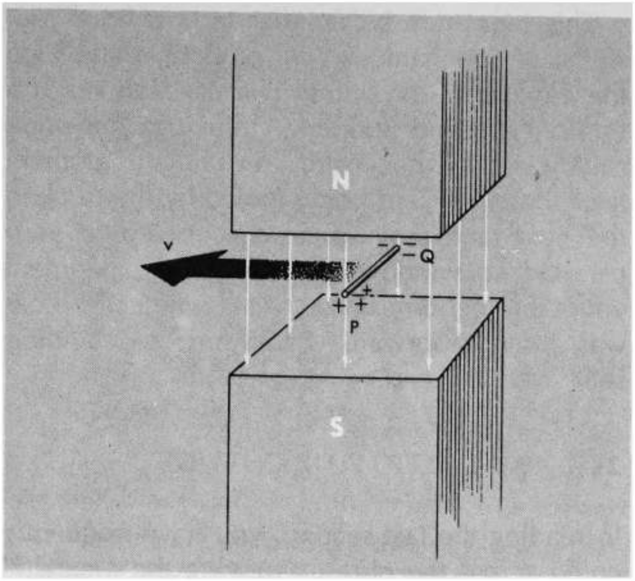
29–1 INDUCED CURRENT

After Oersted's discovery that an electric current is accompanied by a magnetic field, several people began to look for an "inverse" effect: Can a magnet produce a current?

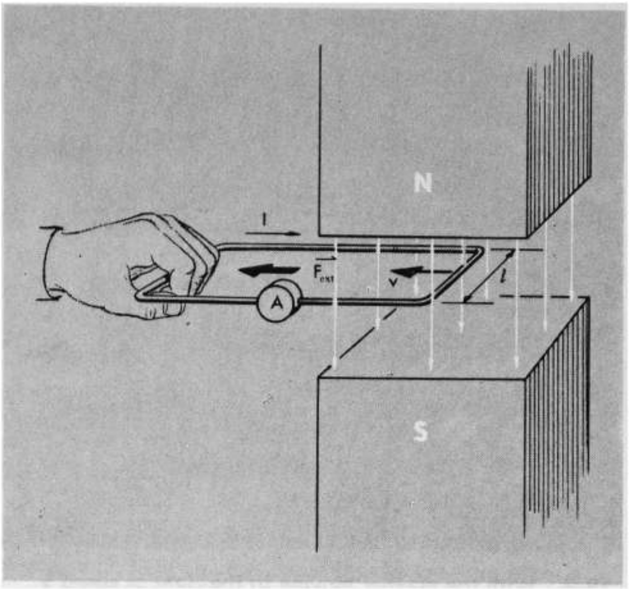
Suppose you wanted to do some experiments to determine whether a magnet can produce a current. How would you proceed? You might wrap some wire around a magnet, connect it to a current meter to complete the circuit, and see what happens. When you found that the meter continues to read zero, you might try a bigger magnet. Disappointed, you try a more sensitive meter. Or, instead of a permanent magnet, you may want to get your magnetic field by passing current through a coil of wire. You would put a piece of soft iron inside the coil, making an electromagnet and thus getting a stronger magnetic field. And if you tried all of these things, you would be treading in the footsteps of that great experimenter, Michael Faraday. But, like Faraday, you wouldn't see the needle of your current meter budge.

Next you might see if what you know about the motion of charged particles in a magnetic field gives you any clues. From Chapter 28 you remember that the force on an electric charge in a magnetic field is proportional to the speed of the charge. So a stationary charge has no force on it, regardless of how strong the magnetic field. Immediately you might conclude: If I want to get a current in a wire in a magnetic field, I have to *move* the wire!

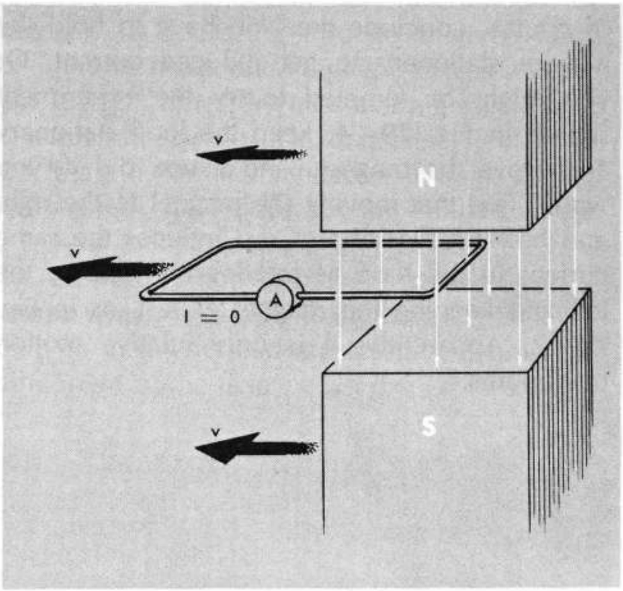
In Fig. 29 – 1, a short, straight wire is being moved to the left across the magnetic field. The right-hand rule tells us the direction of the force on negative charges in the wire. The force is directed toward Q. Positive charges would be pushed toward the near end and negative charges toward the far end of the wire. Of course, you don't expect anything like complete separation of charges, because the electrostatic forces fight back. The positive charges near *P* attract the negative charges near *Q* and repel further positives from their end. This is just what happens in any open circuit when it is connected to a battery: charges build up



29–1 **As the wire moves to the left across the field of the magnet, charges in the wire are pushed toward the two ends.**



29–2 **As the wire loop is pulled across the magnetic field by an external force, current is induced in the loop. Note that part of the loop is in a region of very weak field, well away from the magnet.**



29–3 **When magnet and loop move along together, no current is induced.**

on the ends of the wire. Figure 27 – 2 shows these charges registering on an electrometer.

In Fig. 29 – 2, the circuit has been completed outside the magnetic field, and a meter is inserted to indicate a possible current. Now charge can flow around the circuit. Will the meter register any current?

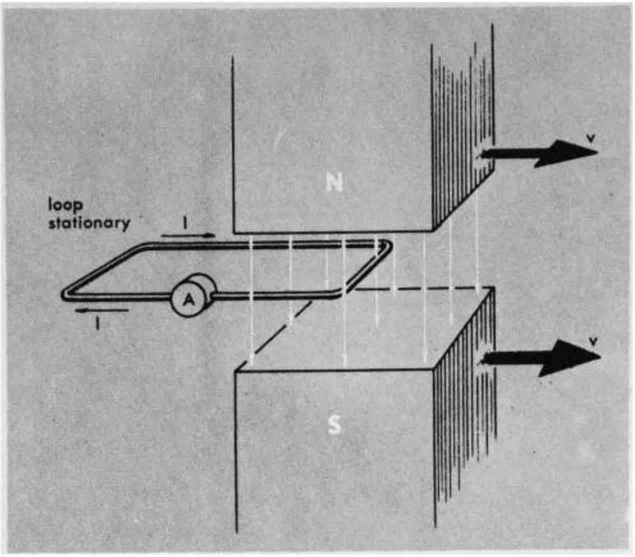
If you were to try this experiment in the laboratory, you would be pretty confident of seeing the meter needle deflect. This time your hopes are based on previous knowledge, not just on a hunch. If you saw nothing, you might again try a bigger magnet, or faster motion of the wire in the field, or a more sensitive meter. And eventually you would detect a current! We say that the motion of the wire loop through the magnetic field has *induced* a current in the loop.

It probably will not surprise you that doubling the velocity of the loop doubles the current; that if you move the loop at the same speed to the right instead of to the left, the current is reversed, but not changed in magnitude; or that turning the magnet upside down does the same thing, just reverses the induced current.

29–2 RELATIVE MOTION

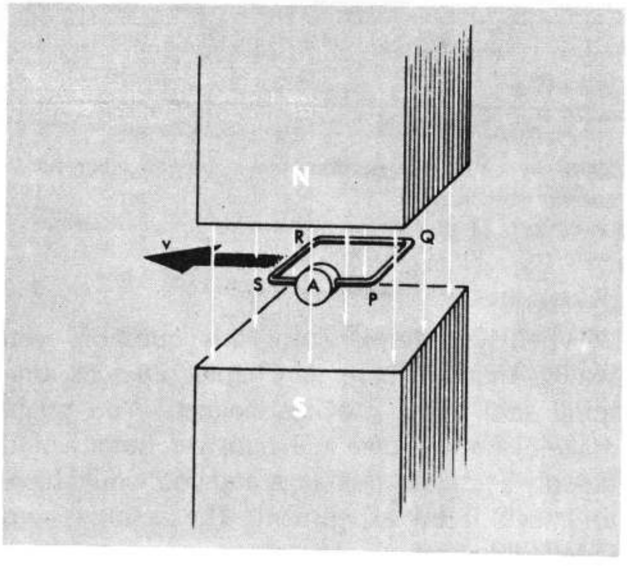
But suppose you moved the wire loop and the magnet together, at the same velocity (Fig. 29 – 3). Would you expect to induce a current in the loop?

When you think about this question, you realize that nowhere in Chapter 28 was anything said about moving magnets. You might make a shrewd guess at the answer, but it would be only a guess at this stage and you would have to check it by experiment. The result is — no induced current.



29–4 With the magnet moving to the right at speed v and the loop stationary, the induced current is the same as with the loop moving to the left at speed v and the magnet stationary. Apparently it is the relative motion that counts.

What do you conclude from this? You might, of course, conclude that you have to hold the magnet stationary to get induced current. Or you might be tempted to try the experiment shown in Fig. 29 – 4: keep the loop stationary and move the magnet. And if you did it, you would find that moving the magnet to the right and holding the loop at rest induces the same current in the loop as moving the loop to the left and keeping the magnet at rest, as in Fig. 29 – 2. Apparently, it is only relative motion that counts.



29–5 When the whole loop is in a uniform magnetic field, the motion induces no current.

This idea may be familiar to you from your study of mechanics. You probably have had the experience of sitting in a car stopped at a traffic light and suddenly thinking you were rolling slowly backward, realizing a moment later that you were being fooled by the truck in the next lane moving slowly forward. If your car had no windows, is there any experiment you could perform inside to tell whether the car was moving forward, backward, or standing still?

29–3 MAGNETIC FLUX CHANGE

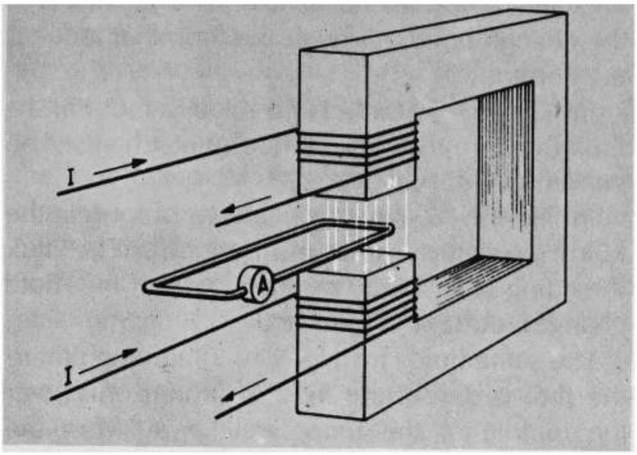
In reading the last section, you have frequently come across the phrase "moving the magnet." How can we say the same thing in the language of magnetic field, without referring to the magnet which is the source of the field? Let us try some more experiments with induced currents and look for hints. Suppose we repeated the experiment of moving a loop in a magnetic field using a very small loop, small enough so that it stays in a region of uniform field throughout the motion (Fig. 29 – 5). It is only when part of the loop gets near the edge of the magnet, where the field gets weaker, that we see the meter needle budge. Is this a hint? Is it that the magnetic field has to be *changing* at some part of the circuit?

We can test that hypothesis without moving either the circuit or the magnet. Using an electromagnet (Fig. 29 – 6), we can change the magnetic field by changing the current in the coils. When we throw the switch to start the current in the coils, a current is briefly induced in the loop. And when we open the switch to stop the current in the coils, a current is induced again, but in the opposite direction.* It looks as if the hint that induced current is related to changing magnetic field has put us on the right track.

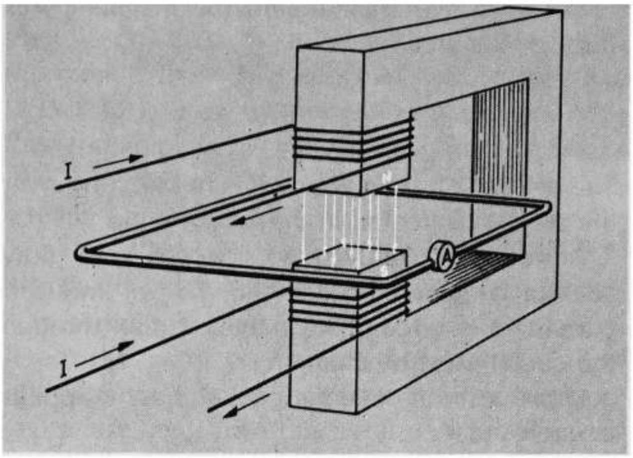
A slight modification of this last setup is shown in Fig. 29 – 7. Here we use a loop large enough so that no part of the wire is between the poles of the electromagnet. All the wire is in a region in which the magnetic field is prac-

---

*Historically, it was through this type of experiment that induced currents were discovered in 1831, by Michael Faraday in England and, independently, by Joseph Henry, who was teaching in a boys' school in Albany, New York.



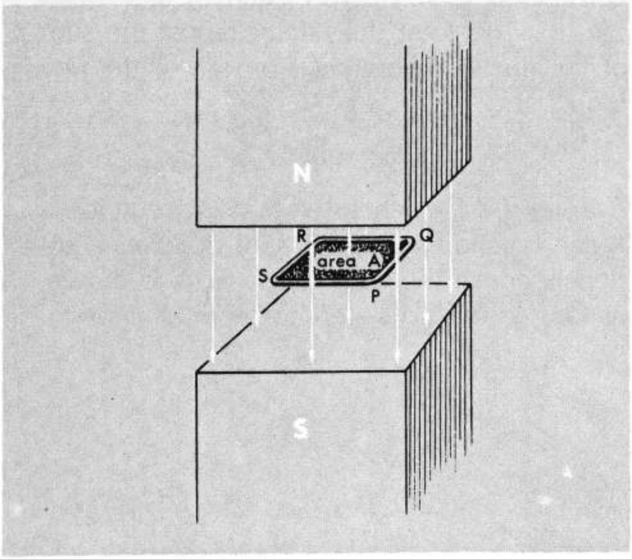
29–6 Current can be induced in a stationary loop by changing the magnetic field of the electromagnet.



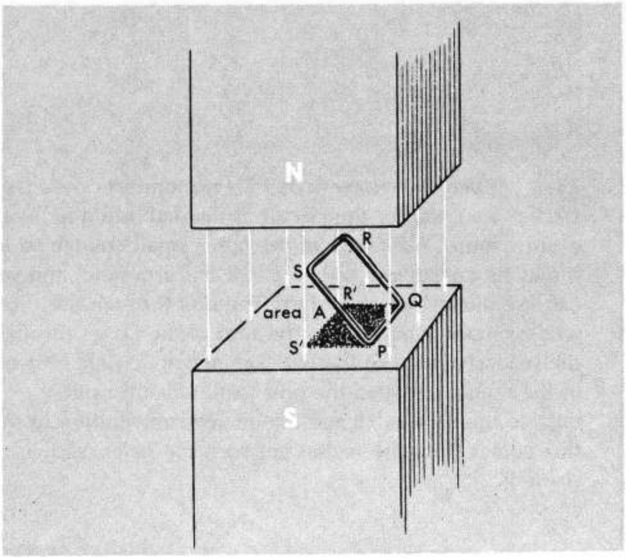
29–7 Current can be induced in a stationary loop when the magnetic field changes, even though the wire itself is in a region of very weak field.

tically zero. But when we throw the switch, the needle of the current meter jumps. So the induced current does indeed seem to depend on a changing magnetic field, but it does not require that the wire itself be in the strong part of the field. What does seem to be important is that we have a change in the "amount" of magnetic field passing through the closed loop of the circuit.

How can we describe the "amount" of magnetic field passing through the circuit loop in a way that applies to all the situations we have discussed (Fig. 29 – 2 through 29 – 7)? Certainly, the magnetic field strength, B, is one important measure of the "amount" of magnetic field, but this only describes the strength at each point. The field is spread out over an area, and we need to take account of the total magnetic field passing through the area enclosed by the loop of the circuit. We shall define a new quantity, the *magnetic flux* ϕ_B, to measure the amount of magnetic field passing through a given area. For uniform fields, the magnetic flux is just the field strength times the area as long as the area is perpendicular to the magnetic field [Fig. 29 – 8 (a)]. If the area is tilted at an angle so that the magnetic field does not run straight across it, less of the magnetic field passes through it [Fig. 29 – 8 (b)]. For example, when we look at an area which is parallel to the lines of magnetic field, no magnetic field crosses it. It is only the component of the field perpendicular to a given



29–8(a) The magnetic flux ϕ_B through the circuit PQRS is the product of the field B times the area A of the circuit: $\phi_B = BA$.



29–8(b) When the circuit PQRS is oblique to the magnetic field, it intercepts less magnetic flux. The effective area A is now the shaded area PQR′S′.

area that carries magnetic field through that area. Thus, we shall define the magnetic flux through the area by

$$\phi_B = B_{\perp}A$$

The general case, in which B is not uniform over the area A, is treated in the box.

Armed with this new concept, we now attempt a general statement: To get induced current in a circuit, the magnetic flux through the circuit must be changing.

Let us test this statement to see how it applies in each situation:

In Fig. 29 – 2, part of the loop is outside the field. As the loop moves to the left, the effective area of loop in the magnetic field decreases, so the loop threads less and less magnetic flux, and current is induced.

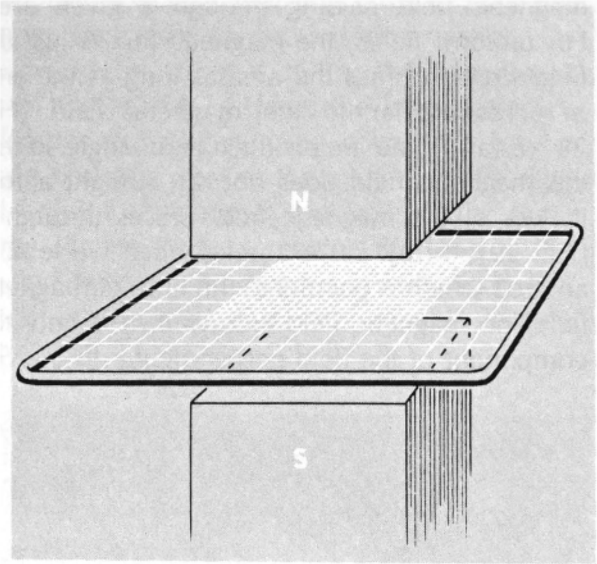
The analysis for Fig. 29 – 4 is the same as for Fig. 29 – 2. It does not matter who moves; it is the change of flux through the loop that induces the current.

In Fig. 29 – 3, there is no relative motion, the flux through the circuit does not change; no current is induced.

In Fig. 29 – 6, the effective area of loop in the field stays the same (no motion). The flux threading the loop changes when the field changes; current is induced.

The same holds for Fig. 29 – 7, although there the flux is dense (the field is strong) only near the middle of the loop, well away from the wire. If this loop is placed with some of the wire in the strong part of the field, less flux is enclosed by the loop, so a given change of current in the coils gives less flux change through the loop and less induced current. To see that

Magnetic Flux



29–9 When the magnetic field is not uniform over the circuit, we consider the circuit as divided into a network of little loops. Each loop of the net is small enough so that it may be considered as being in a uniform field, and we can use our definition of flux. The flux through the whole circuit is defined as the sum of the fluxes through the separate loops of the net. The magnetic field is strong in the region between the poles and falls off rapidly outside this region. Hence, the major contribution to the flux comes from the region between the poles (shown shaded).

The definition $\phi_B = B_{\perp}A$ of the magnetic flux through a circuit loop is incomplete when the $\vec{\mathbf{B}}$ field is not uniform over the area of the loop. For the nonuniform case, it would make sense to use a suitable average of $B_{\perp}$ over the area enclosed by the circuit loop. To find such an average, we shall cover this enclosed area by a network of tiny loops of area ΔA, chosen so small that $B_{\perp}$ can be considered uniform over each of them (Fig. 29 – 9). The total flux, then, is just the sum of the flux contributions from all of the tiny loops, or

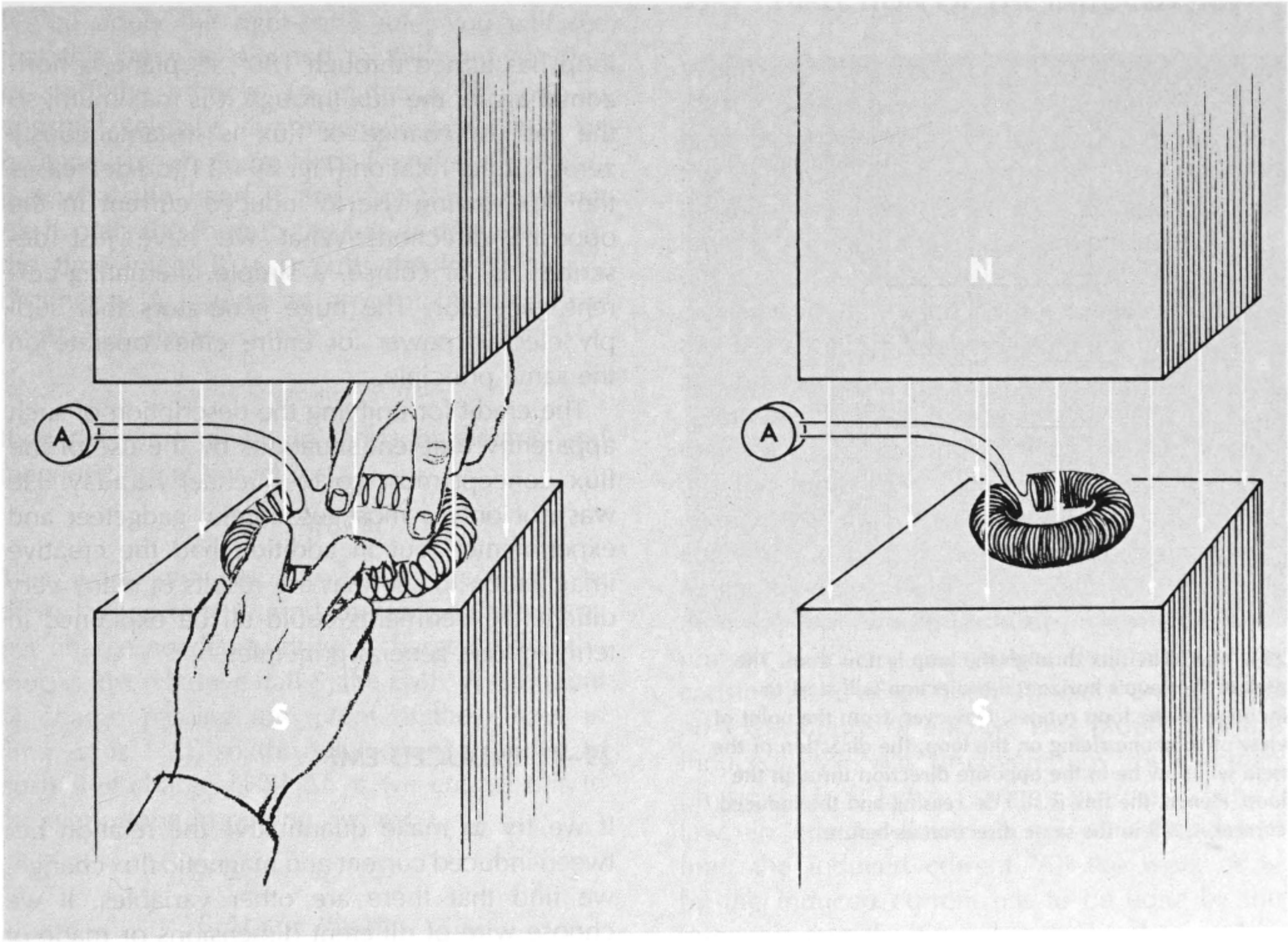
$$\phi_B = \text{sum } (B_{\perp}\Delta A)$$

over all the ΔA's enclosed by the circuit loop. It can be shown that this sum of fluxes is independent of the way the network is drawn, as long as the ΔA's are chosen small enough.

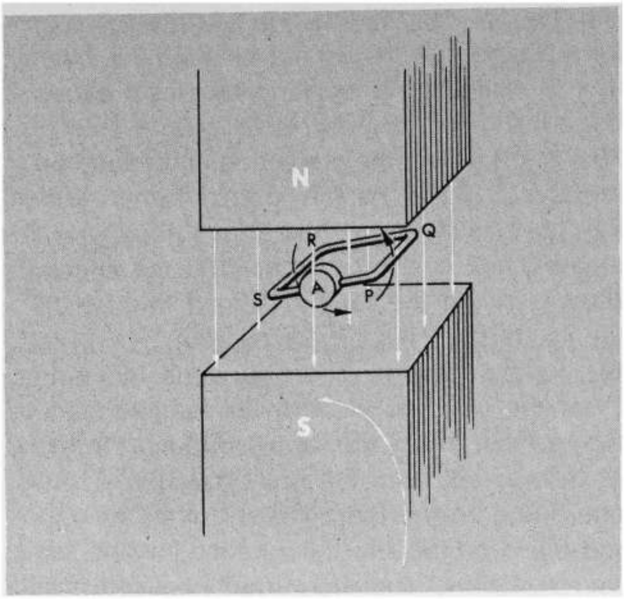
less flux passes through the loop in the displaced position, it may be helpful to think of the electromagnet as providing a given amount of flux for a given coil current. Then, in the displaced position, some of the flux which was passing through the loop now lies outside it. In fact, if the loop is moved over by half its own width, so that one wire goes right down the middle of the face of the electromagnet, it then encloses only half the flux it did before. (Incidentally, if the electromagnet is left on during the displacement, you will see the needle of the current meter twitch.) Now let us look at some new situations:

In Fig. 29 – 10, the field is uniform over a circuit made of a spring. As the circuit is allowed to shrink, the flux through it decreases and current is induced.

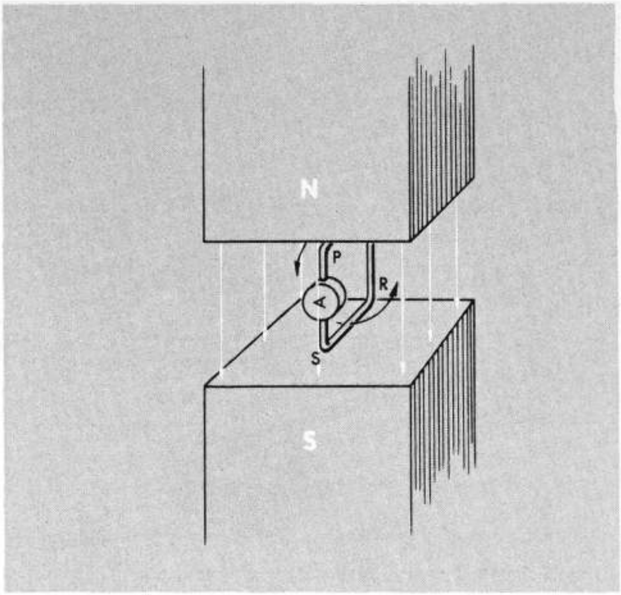
In Fig. 29 – 11 (a), the loop is entirely in a uniform magnetic field, just as in Fig. 29 – 5, but the flux is changed by rotating the loop about a horizontal axis. The effective area A [see Fig. 29 – 8 (b)] is now the projection of the loop on a horizontal plane. As the loop is turned out of the horizontal plane, the flux through it decreases, and current is induced in the direction shown. When the loop is turned through 90°, the flux through it drops to zero [Fig. 29 – 11 (b)]. On further rotation, the flux starts increasing. From the point of view of the magnet (and of the reader), this is flux *down* through the loop, as before; but from the point of view of someone riding on the loop, this is flux in the opposite direction. So the flux change has the same sign as before, and the current induced is still in the same direction in the loop. But when the



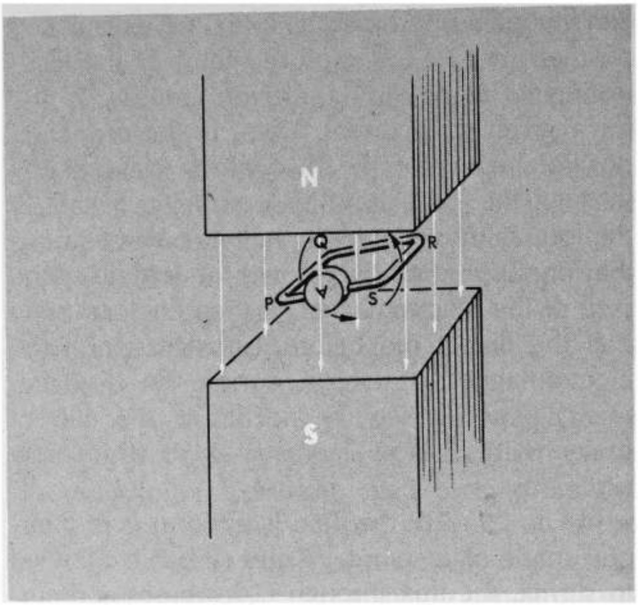
29–10 Current is induced by changing the area of the circuit. Here an extended spring is allowed to contract to a smaller area.



29–11(a) **Current is induced by changing the orientation of the circuit with respect to the magnetic field. As the loop rotates, the area of its projection on the horizontal plane changes.**



29–11(b) **The flux through the loop is now zero. The area of the loop's horizontal projection will start to increase as the loop rotates. However, from the point of view of someone riding on the loop, the direction of the field will now be in the opposite direction through the loop. Hence, the flux is still decreasing and the induced current is still in the same direction as before.**



29–11(c) **The situation is now the same as in (a) except that the loop now is upside down. As viewed from the outside, the current will flow in the same direction as in (a), but the ammeter will register a current in the opposite direction.**

loop has turned through 180°, its plane is horizontal again, the flux through it is maximum, so the rate of change of flux is instantaneously zero. Further rotation [Fig. 29 – 11 (c)] decreases the flux, giving rise to induced current in the opposite direction. What we have just described is, of course, a simple alternating-current generator. The huge generators that supply electric power for entire cities operate on the same principle.

The credit for unifying the description of such apparently different situations by the use of the flux concept must go to Michael Faraday. He was not only a most resourceful gadgeteer and experimenter, but in addition had the creative imagination to see that the results of many very different experiments could all be explained in terms of one general principle.

29–4 INDUCED EMF

If we try to make quantitative the relation between induced current and magnetic flux change, we find that there are other variables. If we choose wire of different dimensions or made of a different metal, the current also is different. Break the wire, and the current stops. The behavior of the current is just the same as if there were a battery in the circuit instead of a changing magnetic flux through the circuit. With a break in a circuit supplied by a battery, a current meter reads zero, but an electrometer indicates a potential difference across the break equal to the EMF of the battery. This leads us to try to analyze the situation of Fig. 29 – 2 to see if a given motion corresponds, not to a definite current, but to a definite EMF. We shall try to relate this EMF to the rate of change of magnetic flux. The relation we obtain, put to the test of experiment in "new" situations, will be found to hold generally, regardless of how the flux change is caused.

In Fig. 29 – 2, suppose that uniform motion of the loop with velocity $\vec{\mathbf{v}}$ to the left across the magnetic field $\vec{\mathbf{B}}$ induces a current I in the loop as shown. Then, as you know from Section 28 – 4, the force on the length l of wire due to the magnetic field is

$$F_{\text{magnetic}} = IlB$$

If you apply the right-hand rule, you will see that this force is directed to the right, tending to stop the motion. To maintain the loop in constant velocity, this magnetic force must be balanced by an equal force F to the left; that is just what the hand in the sketch is supplying. As it pulls the loop to the left, it does work. In the time interval Δt, it pulls the loop along a distance $v\ \Delta t$, doing an amount of work $W = Fv\ \Delta t$. But, since

$$F = -F_{\text{magnetic}} = -IlB$$

then the work done, in terms of electric and magnetic quantities, is just

$$W = -IlBv\Delta t$$

This work provides the energy necessary to drive the current around the loop. The energy per charge needed to drive a charge around the loop is just what we call $\mathcal{E}$, the EMF. The amount of charge passing any point of the circuit in time Δt is $I\ \Delta t$, so the total energy needed to push this charge is $\mathcal{E}I\ \Delta t$. If we equate this to the work done in pulling, we get

$$\mathcal{E}I\ \Delta t = -Il\,Bv\Delta t$$

or $$\mathcal{E}\ \Delta t = -lBv\Delta \mathbf{t}$$

The disappearance of I from the equation is important, since it tells us that the length of the wire, the field strength, and the velocity are related directly to the induced EMF, rather than to the current. Why have we not eliminated the Δt as well? Because the length l moving a distance $v\ \Delta t$ sweeps out an area $lv\Delta t$, the magnetic flux change is

$$\Delta\phi_B = Blv\Delta t$$

We substitute this into the right-hand side of the previous equation and solve for the induced EMF

$$\mathcal{E} = -\frac{\Delta\phi_B}{\Delta t}$$

We see that the induced EMF is proportional to the rate of change of the magnetic flux. That the proportionality constant turns out to be unity is due to the choice of the unit of magnetic field strength made in Section 28 – 4.

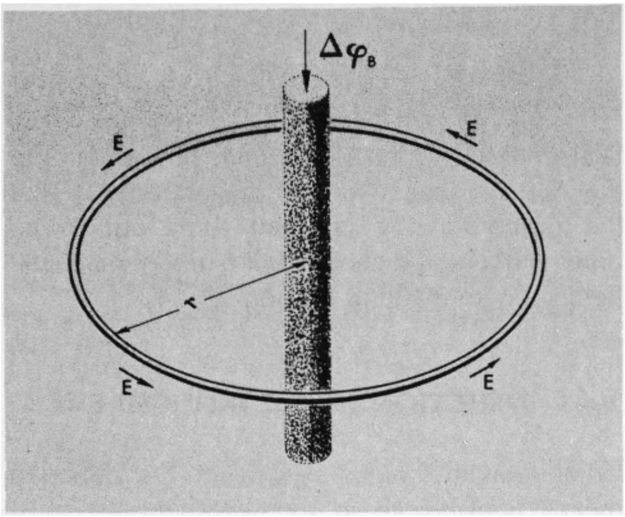
29–5 DIRECTION OF THE INDUCED EMF

What does the minus sign in the last equation mean? Somewhere in the argument we must have assigned a meaning to positive or negative flux without saying so explicitly. In the identification of $+Blv\Delta t$ as the flux change, we must have introduced a sign convention for flux. But rather than analyze what we mean by positive or negative B, v, and Δt separately, let us use the conservation-of-energy principle to arrive at an understanding of the negative sign.

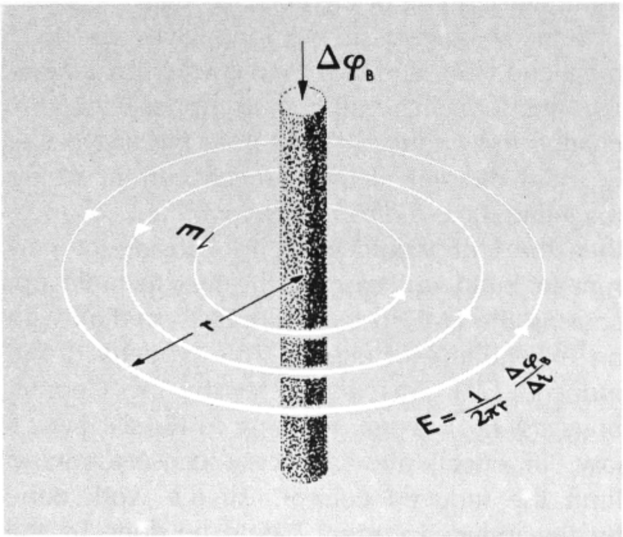
Induced current in the loop gives rise to a magnetic field, and thus to magnetic flux threading the loop. This flux must oppose the flux change that induced the EMF in the first place. For if it did not, if the induced current set up magnetic flux *aiding* the overall flux change, then the EMF would build up, causing the current to build up, causing the flux to build up, causing the EMF to build up, etc.—and all with no expenditure of energy. This property of the induced EMF was discovered by the German physicist H. F. Lenz, working in Russia. Lenz's law, in effect, allows energy conservation to limit the induced current. All the work done by the induced current has to be done by the force causing the flux change: by the hand in Fig. 29 – 2, by the turbine in the power-generating station that supplies your town's electricity.

29–6 ELECTRIC FIELDS AROUND CHANGING MAGNETIC FLUXES

We can get a fuller understanding of the electromagnetic induction equation of Section 29 – 4 if we rewrite it, this time using "field" language rather than EMF. We owe this field description to James Clerk Maxwell, who also introduced another field relation (to be discussed in Section 29 – 7) and was then able to predict electromagnetic waves for the first time.



29–12 When the flux changes through the shaded area, the magnitude of the electric field that circles around it must be the same everywhere along the wire. Therefore the energy per elementary charge received by a charge pushed all the way around the wire is 2πrE.



29–13 The induced electric field around a changing magnetic flux runs around in circles. Its magnitude varies inversely with the distance.

We begin this train of thought with an observation: magnetic field lines around currents always close. They do not start or stop anywhere. In fact, the field lines due to current in a very long, straight wire are perfect circles around the wire. On the other hand, the electric field lines described in Chapter 25 start on positive and end on negative charges. Is it possible for electric field lines to form closed loops, too?

To answer this, let us consider a circular wire loop of radius r with a changing magnetic flux along its axis (Fig. 29 – 12). The law of induction tells us that an EMF is induced in the loop whenever there is a change of flux through the loop. There will be an induced current in the wire, which means that charges are being driven around the loop. There must be a net force on these charges to drive them around the circuit—to supply the energy used in heating the wire, for example. This force, measured per elementary charge, is just the electric field we are looking for. In this nice, symmetrical example, there is nothing to distinguish one part of the circular loop from any other, so the electric field must have the same magnitude everywhere in the loop and must point along the wire.

We can find the magnitude of $\vec{\mathbf{E}}$ from the law of induction. The EMF was defined in Chapter 25 as the work per elementary charge, in this case in going all the way around the circuit, a distance $2\pi r$. The work per elementary charge is then related to the force per elementary charge by the relation $\mathcal{E} = 2\pi rE$. In terms of the rate of change of flux, we get

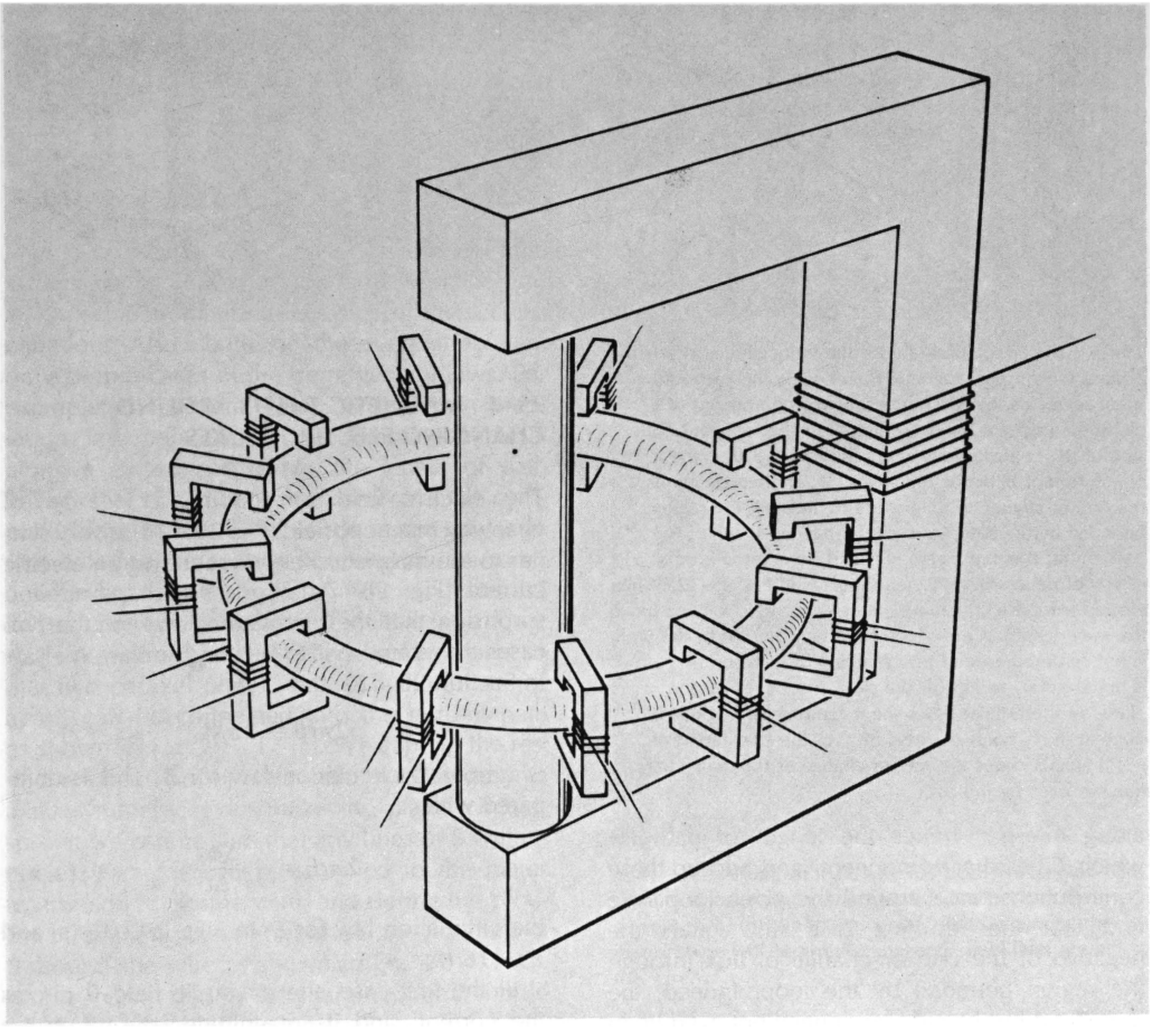
$$2\pi rE = -\frac{\Delta\phi_B}{\Delta t}$$

We see that the electric field does have constant magnitude around a circle of radius r, just as we expected. But we also see that if we have a wire loop of greater radius, the E field at that greater radius must be proportionately smaller, since the product of the field times the circumference must always equal the negative of the rate of change of flux.

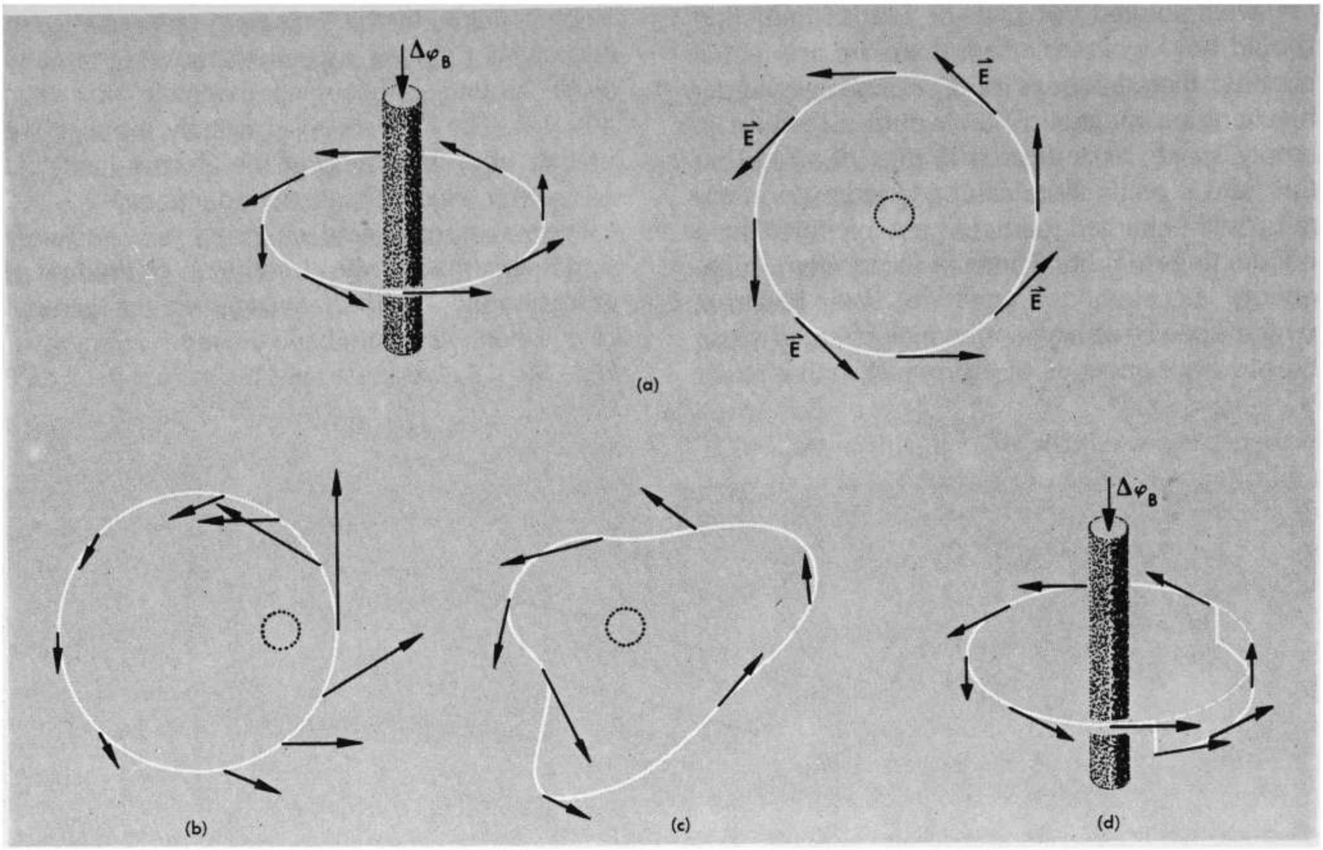
But in Chapter 25 the electric field was defined to describe the force per charge regardless of whether a charge is really there or not.

Maxwell pointed out that the law of induction should hold whether or not there are any actual circuits, that the lines of *E* around a cylinder of changing magnetic flux would also exist in empty space, as indicated in Fig. 29 – 13. That this field is really there can be tested experimentally with charged particles not confined to a conducting circuit. There is a modern high-energy accelerating machine, the betatron, which speeds up the individual charged particles to high energies by using just such circular electric fields. In this machine, electrons in an evacuated ring are caused to move faster and faster around a changing magnetic flux (Fig. 29 – 14). They speed up in exactly the way we predict from the action of the electric field that arises from electromagnetic induction.

We have considered only a special and very simple example of the description of the law of induction in "field" language. In the general case, when we do not have the nice symmetry of Fig. 29 – 12, we must take the component of $\vec{\mathbf{E}}$



29–14 A schematic diagram of a betatron. The changing flux in the center produces an electric field which accelerates electrons. The electrons are made to move in circular paths by the outer ring of magnets. We correctly predict the paths from the local magnetic field for deflection and from the induced electric field giving the acceleration along the path.



29–15 The circulation law for the induced electric field around a changing magnetic flux. Divide the path into small segments. For each one take the component of $\vec{E}$ along the segment times the length of that segment. The sum of the contributions along the entire path is the circulation of $\vec{E}$, or the induced EMF, and equals minus the rate of change of magnetic flux through the region bounded by the path, no matter what the shape of the path. In (a) the path is circular and is centered around a region of decreasing magnetic flux, as in Fig. 29–12. When viewed from directly above, the electric field vectors have the same length at every point on the path. In (b) the path is not centered around the changing flux, in (c) the path is not circular, and in (d) the path is not even in one plane. In these three cases the magnitude of $\vec{E}$ is not the same at every point on the path. Yet the circulation of $\vec{E}$ still equals minus the rate of change of the magnetic flux enclosed by the path.

along the path times the length of path for which $\vec{E}$ has that component, and add up these contributions once around the given loop. We find that the resulting EMF still equals the negative of the rate of change of flux through the region bounded by the loop. Indeed, the loop need not be circular, it need not even lie in a plane; the law of induction still holds (Fig. 29 – 15). It is one of the fundamental laws of physics. It is often called a "circulation law" for electric fields.

29–7 MAGNETIC FIELDS AROUND CHANGING ELECTRIC FLUXES

The electric field surrounding a cylinder of changing magnetic field (Fig. 29 – 13) is very similar to the magnetic field surrounding an electric current [Fig. 28 – 7 (a)]. It is perhaps not too surprising that the circulation laws for the two cases also came out to be very similar. We saw in Section 28 – 10 that

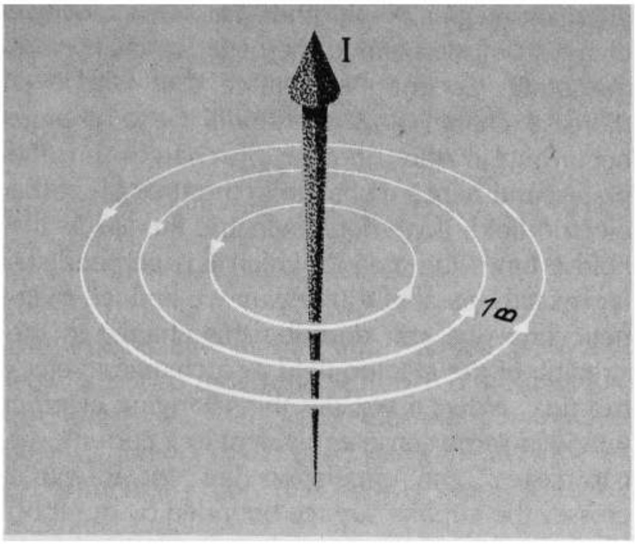
$$2\pi rB = 2\pi KI$$

is Ampère's circulation law for $\vec{B}$. This is compared with

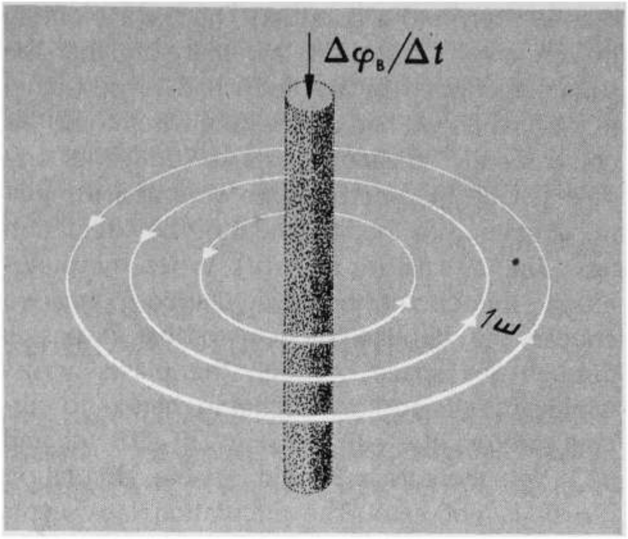
$$2\pi rE = -\frac{\Delta\phi_B}{\Delta t}$$

the circulation law for $\vec{E}$, in Figs. 29 – 16 (a) and 29 – 16 (b).

In the first case, the magnetic field $\vec{B}$ circles the current, and its magnitude falls off as $1/r$ if the current remains constant; in the second case, the electric field $\vec{E}$ circles the changing magnetic flux, and its magnitude also falls off as $1/r$ if the rate of change of magnetic flux re-



29–16(a) **The magnetic field of an electric current circles the current, falling off in magnitude as 1/r.**



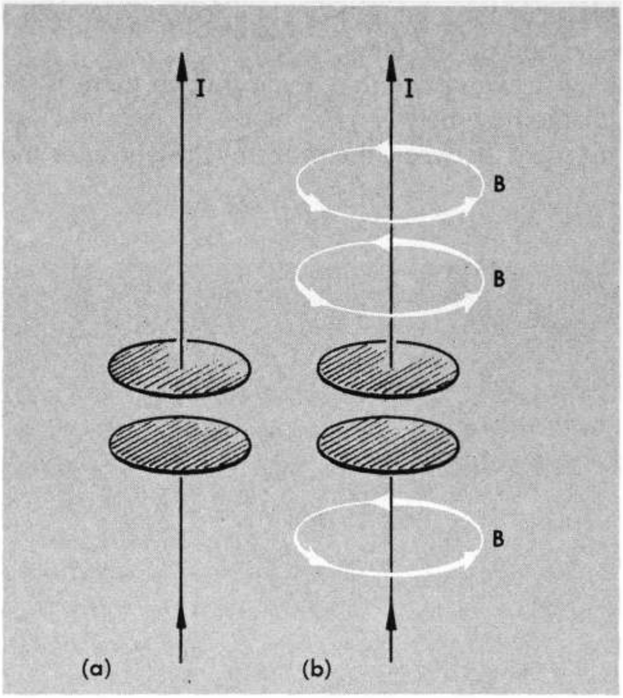
29–16(b) **Similarly the electric field around a changing magnetic flux circles the flux and falls off as 1/r far from the region of strong magnetic field.**

mains constant. Only the direction of the arrows along the axes of the loops are different, as expected from the minus sign of the second equation. Also, both of these equations are only simple cases of the more general laws. For example, the general form of Ampère's circulation law tells us that the component of $\vec{\mathbf{B}}$ along a closed path times the length of path having that component, summed over the complete loop, is equal to $2\pi K$ times the total current crossing the surface bounded by the loop.

But what happens if the current is not steady? Let us consider the magnetic field near a long, straight wire which has been broken open so that two parallel plates, spaced a short distance apart, can be connected into the current path as shown in Fig. 29 – 17 (a). We can put the rest of the circuit so far away that we can assume axial symmetry, even make the plates circular so that we can be sure that any lines of B will be circles. Charges will be supplied to the plates by the current in the wire, and during the process of charging there should be circular lines of B around the wire as shown in Fig. 29 – 17 (b). At any instant during the charging, this B should satisfy Ampère's circulation law:

$$2\pi rB = 2\pi KI$$

with I the instantaneous current through the surface bounded by the circular line of B. But



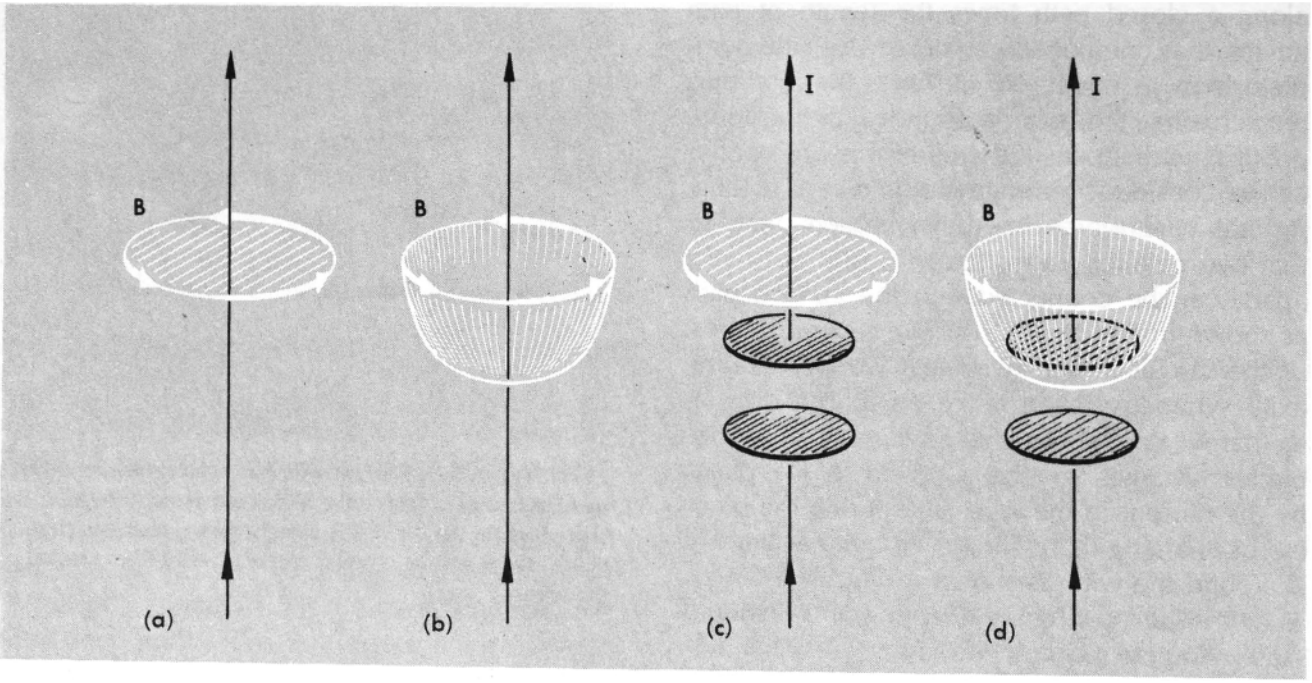
29–17 **(a) A long straight wire has been broken and two parallel circular plates have been connected into the current path. (b) While the current flows, charging the plates, there will be circular magnetic field lines around the wire.**

here we run into a paradox! There was nothing in Ampère's circulation law that said that the surface bounded by the loop had to be plane. In fact, when we had a very long wire alone, we could readily consider a cuplike surface (or any other shape bounded by the loop) and it made no difference to the statement of the circulation law [Fig. 29 – 18 (a) and (b)]. When the plates are in the circuit, however, suppose that we choose the cuplike surface so that it passes through the space between the plates across which there is no flow of charge, hence no current through the surface [Fig. 29 – 18 (c) and (d)]. Yet, there is a $\vec{\mathbf{B}}$ field, so the circulation ($2\pi rB$) is not zero. The circulation law seems to have failed. We have a genuine paradox.

To resolve this problem we must do one of two things: either we must say that the circulation law for $\vec{\mathbf{B}}$ holds only for steady currents, or we must amend the law so that it holds also for this case where charges build up at some place in the circuit. We shall follow Maxwell's lead and do the latter.

First, we notice that, even though there is no charge flow between the plates, there is an electric field between them at any instant after the charging begins, with lines from the positive charges on one plate to the negative charges on the other. Second, we notice that whenever there is charge flow there will be a change occurring in the number of charges on the plates, and hence a change in the strength of the electric field. If we define electric flux ϕ_E as the field E times the area to which $\vec{\mathbf{E}}$ is perpendicular, exactly parallel to the way we defined magnetic flux, we can describe this change in the strength of the electric field as a change in electric flux. Now, if we say this changing electric flux is in some sense equivalent to a current, we can "save" the circulation law for $\vec{\mathbf{B}}$. What crosses the cuplike surface bounded by our loop [Fig. 29 – 18 (d)] is not a real current but a changing electric flux.

In the box it is shown how the change in electric flux is related to the change in the charge on the plates. We see there that the total current needed to make Ampère's circulation law correct when the currents are not steady is the sum of the true current, I, and the "equivalent current," $\frac{1}{4\pi k}\,\frac{\Delta\phi_{\mathrm{E}}}{\Delta t}$, where k is the constant in



29–18 (a) and (b) Nothing in Ampére's circulation law (applied here to the field of a long, straight wire carrying a current) said that the surface bounded by the loop had to be plane. (c) and (d) Applying the same reasoning when parallel plates are in the circuit, however, leads to a genuine paradox.

Coulomb's law. For the cuplike surface passing between the plates, the true current is zero and the "equivalent current," due to the changing electric flux, is all-important. There,

$$2\pi r B = \frac{2\pi K}{4\pi k}\frac{\Delta\phi_E}{\Delta t}$$

$$= \frac{1}{9\times 10^{16}\ \text{m}^2/\text{sec}^2}\frac{\Delta\phi_E}{\Delta t}$$

The importance of this term, added by Maxwell, is not so much that it resolved the paradox but that it led him to predict electromagnetic waves, and to say, many years in advance of experimental proof, that light is simply electromagnetic radiation.

The fact that a changing electric flux produces a magnetic field has always been very difficult to show experimentally. Even today, the existence of electromagnetic waves, behaving just as Maxwell predicted, is the most convincing proof of the equivalence between current and the changing electric flux.

29–8 THE MECHANISM OF ELECTROMAGNETIC RADIATION

The two circulation laws enable us to derive the important properties of electromagnetic waves in empty space. We must first rewrite the laws to include loops which are made of straight-line

There are many ways of seeing exactly how the electric flux is related to the charge. One of the easiest is to consider a small body with a charge Q on it, which we can surround, in imagination, with a sphere of radius r. The electric field which is directed straight outward along the radii is of strength

$$E = kQ/r^2$$

The area of the sphere is $4\pi r^2$. Consequently, the total outward flux (field times area, which is in this case everywhere perpendicular to the field) is

$$\phi_E = 4\pi r^2 E = 4\pi k Q$$

The flux is thus independent of the radius; it is the same for all spheres surrounding the charge Q. Hence the rate of change of electric flux is

$$\frac{\Delta\phi_E}{\Delta t} = 4\pi k\frac{\Delta Q}{\Delta t}$$

But $\Delta Q/\Delta t$ is just the current flowing onto the body to charge it. Since

$$\frac{\Delta Q}{\Delta t} = \frac{1}{4\pi k}\frac{\Delta\phi_E}{\Delta t}$$

the quantity $\frac{1}{4\pi k}\frac{\Delta\phi_E}{\Delta t}$ must have the magnitude of the charging current at any instant. It is a sort of "equivalent current" in a region where there is not a true current but a changing electric field instead.

We can use this "equivalent current" to amend the circulation law for $\vec{\mathbf{B}}$ in our problem with the parallel plates connected to the very long wire. To the true current we add the "equivalent current" term and obtain

$$2\pi r B = 2\pi K\left(I + \frac{\Delta Q}{\Delta t}\right)$$

$$= 2\pi K\left(I + \frac{1}{4\pi k}\frac{\Delta\phi_E}{\Delta t}\right)$$

In Section 24 – 7 we found that $k = 2.3 \times 10^{-28}$ newton-meter²/(elem. ch.)². Since we now know that one elementary charge is the same as 1.6×10^{-19} coulombs, we can readily determine that in practical units

$$k = 2.3\times 10^{-28}/(1.6\times 10^{-19})^2$$
$$= 9\times 10^{9}\ \text{newton-meters}^2/\text{coul}^2$$

From Section 28 – 11 we know that

$$K = 2\times 10^{-7}\ \text{newton/amp}^2$$

Therefore the contribution to the circulation from the "equivalent current" is

$$\frac{2\pi K}{4\pi k}\frac{\Delta\phi_E}{\Delta t} = \frac{4\pi\times 10^{-7}\ \text{nt/amp}^2}{4\pi(9\times 10^{9}\ \text{nt-m}^2/\text{coul}^2)}\frac{\Delta\phi_E}{\Delta t}$$

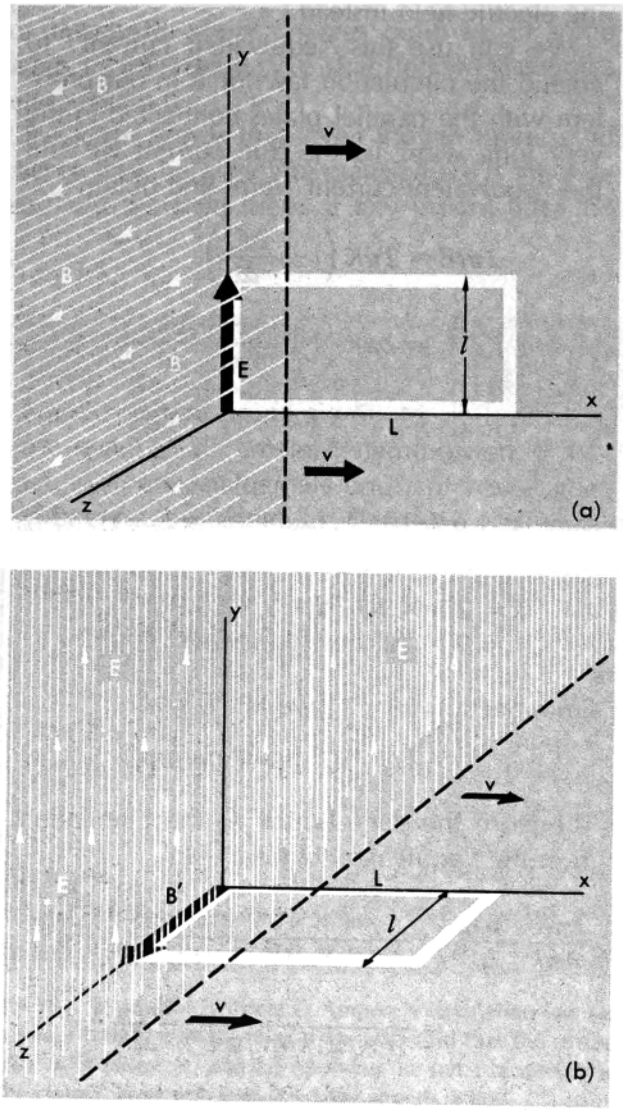
$$= \frac{1}{9\times 10^{16}\ \text{m}^2/\text{sec}^2}\frac{\Delta\phi_E}{\Delta t}$$

segments instead of symmetrical circular paths. We need the component of each field in the direction of the path at that point; we multiply by the length of path for which the field component has this particular value and then sum up the result over all segments of the loop. Thus,

$$\mathcal{E} = \Sigma E_\ell \ell = -\frac{\Delta\phi_B}{\Delta t}$$

$$\Sigma B_\ell \ell = \frac{1}{9 \times 10^{16}\,\frac{\text{m}^2}{\text{sec}^2}}\,\frac{\Delta\phi_E}{\Delta t}$$

where the capital sigma (Σ) means "sum of" around the whole loop. (We have left out the ordinary current term I because there are no moving charges in the part of space we shall be considering—in fact, no stationary charges either, since there is no material substance at all.) These are the two laws which Maxwell combined to predict electromagnetic waves. We can follow his argument physically if we *assume* that a changing field exists in space, and examine the circulation laws to see what would happen.



Let us consider a magnetic field $\vec{\mathbf{B}}$, uniform where it exists but with a sharp front face or edge, so that $B = 0$ to the right of this face and B = constant to the left of it. Assume that this boundary between "field" and "no field" moves through space with a velocity which we call $\vec{\mathbf{v}}$. Figure 29 – 19 (a) indicates such a field, pointing toward us in what is shown as the z direction, with its boundary moving to the right, in the x direction. Now think of a rectangular loop lying in the x-y plane, at right angles to $\vec{\mathbf{B}}$; its y dimension is ℓ, and its length L is greater than the distance the boundary will move in the time Δt for which we consider the motion. The figure indicates the beginning of the interval Δt, with one leg of the loop lying just inside of the field region. Then in time Δt the flux through the loop will increase by

$$\Delta\phi_B = Bv\ell\,\Delta t$$

This expression looks very familiar; it is, in fact, much like an equation near the end of Section 29 – 4. It is true that there we were dealing with a real circuit and here our loop is imaginary, but we have found that the circulation law holds just as well for imaginary loops as for real wire circuits. There is a flux change through the surface bounded by the loop, and therefore an EMF. If we look to see what contributes to the sum of all the $E_\ell \ell$'s, we notice that the right end of the loop is far away, and the contributions from the two sides parallel to x would cancel each other as one traverses the whole loop. There *must* be an $\vec{\mathbf{E}}$ somewhere to satisfy the circulation law, and an $\vec{\mathbf{E}}$ upward in

29–19 (a) A region of space containing magnetic field B = constant to the left, and B = 0 to the right of a sharp boundary. The boundary, a plane parallel to the y-z plane, is assumed to move to the right with a velocity $\vec{\mathbf{v}}$. Magnetic flux through a rectangular loop in the x-y plane increases at a rate Blv, inducing an electric field in the left leg of the loop. (b) The induced electric field causes a changing electric flux in a rectangular loop in the x-z plane at the rate Elv, inducing a magnetic field $\vec{\mathbf{B}}'$ in the same direction as the original $\vec{\mathbf{B}}$.

the y direction along the back segment of the loop will just do it. We can get the magnitude of $\vec{\mathbf{E}}$ from

$$\frac{\Delta\phi_B}{\Delta t} = Bvl = El$$

Therefore, $E = vB$ at the left of the loop. Can you check that $\vec{\mathbf{E}}$ is *up* along y, not down, by Lenz's law?

This electric field, at right angles to $\vec{\mathbf{B}}$, must exist everywhere to the left of the moving edge of $\vec{\mathbf{B}}$, since to apply the circulation law for $\vec{\mathbf{E}}$ we could use any loop which is partly inside and partly outside the region. Thus the field $\vec{\mathbf{B}}$ whose boundary moves with velocity $\vec{\mathbf{v}}$ is accompanied by an electric field at right angles to both $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$, of magnitude $E = vB$. But the circulation law relating B to the change of *electric* flux must also be satisfied. If we consider a rectangular loop in the x-z plane, at right angles to the loop we considered before, there is a change of electric flux through its surface in time Δt, and therefore an induced magnetic field somewhere along the perimeter of the loop [Fig. 29 – 19 (b)]. Using arguments which are exactly similar to those we have just used, we get

$$\Delta\phi_E = Evl\, \Delta t$$

and

$$\frac{\Delta\phi_E}{\Delta t} = Evl$$

If we call $\vec{\mathbf{B}}'$ the magnetic field induced according to the second circulation law, it follows that there must be a field $\vec{\mathbf{B}}'$ everywhere behind the advancing front of $\vec{\mathbf{E}}$, and at right angles to $\vec{\mathbf{E}}$, such that

$$B' = \frac{v^2B}{9 \times 10^{16}\ \text{m}^2/\text{sec}^2}$$

This $\vec{\mathbf{B}}$, which accompanies $\vec{\mathbf{E}}$, is in the same direction as the $\vec{\mathbf{B}}$ with which we began. If we eliminate E from the two equations, we find

$$B' = \frac{v^2B}{9 \times 10^{16}\ \text{m}^2/\text{sec}^2}$$

Thus far, we have said nothing about the magnitude of v. Now we see that if

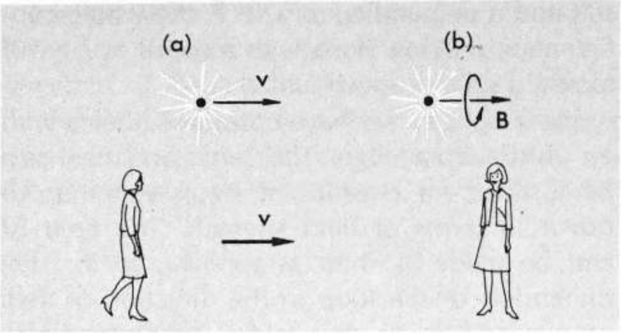
$$v = 3 \times 10^8\ \text{m/sec}$$

then $B' = B$; in short, the field $\vec{\mathbf{B}}$ maintains itself and a perpendicular field $\vec{\mathbf{E}}$, the whole configuration moving along with a speed of 3×10^8 m/sec, a unique speed called c.

For simplicity we have considered fields with an abrupt front edge. The same argument can be justified for a series of steps, either up or down in terms of field strength. The time Δt can be made as short as we like, so that the dimension of the loop in the direction of field propagation may be taken correspondingly small; for every increment (positive or negative) of one field there is a corresponding increment of the other, satisfying the same relation we have shown in detail for a sudden step from zero to a uniform field. Thus, the field changes may be continuous from one point to the next, and can have any arbitrary profile in space; for example, the field may vary in space like a sine function. At any point in such a moving configuration, $E = cB$ at any instant, and the field pattern moves through space with speed c, which is just the speed of light. This conclusion from the circulation laws for electric and magnetic fields led Maxwell to conclude "that light itself (including radiant heat, and other radiations if any) is an electromagnetic disturbance in the form of waves. . . ." In Maxwell's time there was no direct evidence of the electrical character of light, or of any other radiation. More than twenty years after this prediction Heinrich Hertz was able to generate radiation which we now call radio waves, using machinery that was obviously electrical, and to show that the waves do travel with the speed of light. Then Maxwell's theory was brilliantly confirmed. In the next section we shall look briefly at the evidence for the existence of electromagnetic radiation and its electrical nature.

29–9 EVIDENCE FOR ELECTROMAGNETIC RADIATION; THE ELECTROMAGNETIC SPECTRUM

All electric and magnetic fields arise from charges and their motions. When should we expect these fields to radiate away from the electric particles from which they originate? An electric charge at rest has only a radial Coulomb field which falls off as the inverse square of the distance and does not propagate



29–20 **(a) If we walk along beside a moving charge, we see only its Coulomb field. (b) If the charge moves past us, we also see a magnetic field in the same region.**

away from it. A charge in constant-velocity motion is really no different from a charge at rest (Fig. 29 – 20). If we walk alongside it, we see the charge at rest and the field around it is a Coulomb field. If we move past it, or it moves past us, because of the changes of electric flux through any region, we shall also see a magnetic field. But this magnetic field is as much attached to the moving charge as the Coulomb field. So the electric and magnetic fields stay with a charge in constant-velocity motion. They do not radiate away.

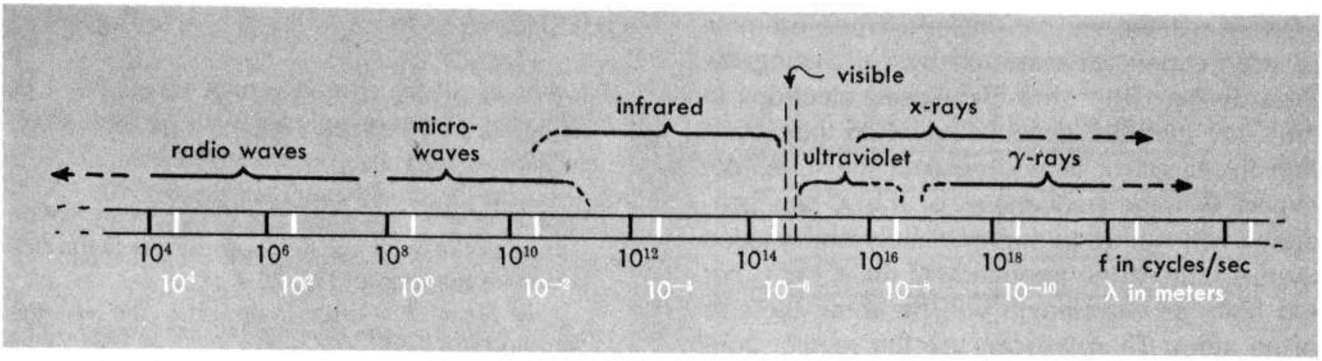
On the other hand, if we accelerate a charge, it cannot stay at rest in any dynamical frame of reference in which Newton's laws and the laws of electromagnetism apply. At that moment, we have our only chance to start a pulse of crossed electric and magnetic fields moving away. To find electromagnetic radiation, then, we should look for radiation from accelerated charges.

Let us check up experimentally on what happens when we accelerate charges. A radio transmitting station pumps charges along an antenna, first in one direction and then in the other direction. They do not move at constant velocity but go back and forth, accelerated first one way and then the other. The radio waves which break away from the antenna travel at 3×10^8 meters per second. Evidently they arise from the accelerated motions of the charges pushed back and forth along the antenna. When we stop high-speed electrons on the target of an X-ray tube, we give them a rapid backward acceleration to bring them to rest. X rays radiate away from the region where the charges are accelerated. On measuring the speed of propagation of X rays, we find that it also is 3×10^8 meters per second.

In a synchrotron—a machine designed to accelerate electrons to high speeds—electrons of tremendous energies are forced around large orbits by deflecting magnetic fields. If we look around the orbit, we see light coming out. Indeed, the energy to which we can bring the electrons is limited in practice by the energy they radiate in light as they are accelerated by the deflecting magnetic forces. We feed in energy from a huge electrical oscillator as rapidly as we can, but at some speed of motion of the electrons it is radiated away as fast as we can supply it. Ordinary visible light is therefore an electromagnetic radiation with its origin in the acceleration of electric particles. The coincidence between the speed of light and the speed of electromagnetic radiation as calculated by Maxwell was no accident. All the radiations we have mentioned travel at 3×10^8 meters per second; all of them are electrical in origin. They are all electromagnetic radiation (Fig. 29 – 21).

The common speed of these radiations and their origin from accelerated charges is not the only evidence we have for their electromagnetic nature. There is other evidence connecting the whole spectrum of radiation from X rays to radio waves. If we accelerate charged particles back and forth with a given frequency, the radiant electric fields that we shake off from them should have this frequency of oscillation. Each radio transmitting station pumps charge back and forth through its antenna at a definite frequency. It must impart this frequency to the radio waves which it sends out. The oscillating electric field in these waves should in turn drive charges back and forth in the receiving antenna at the same frequency. Are those oscillating currents really there? We detect the radio signals by tuning the radio receiver, and we pick up a particular radio station when we have tuned our receiver to exactly the same frequency that the transmitter puts out. The oscillation frequency has been carried from the transmitter to us by the electromagnetic radiation.

With various circuits and antennas we can make transmitters that will send out radio waves of frequencies from about 10^4 cycles per second to about 3×10^{11} cycles per second.



29–21 The electromagnetic spectrum: this is a continuous range of radiation spreading from gamma rays to radio waves. The descriptive names for sections of the spectrum are historical; they merely give a convenient classification according to the source of the radiation. The physical nature of radiation is the same throughout the whole range. In all sections it has the same speed, the same electromagnetic nature, and the only difference from one part of the spectrum to another is in frequency and wavelength. The regions with historical names overlap, but the names still give a hint of the common sources: radio waves and microwaves from electrons moving in conductors; infrared from hot objects; visible light from very hot objects; ultraviolet from arcs and gas discharges; X rays from electrons striking a target; gamma rays from nuclei of radioactive atoms.

All of these electromagnetic waves travel with the speed of light. So the wavelengths—which we can measure by interference—run from about 30 kilometers down to the region of a millimeter. At the long-wavelength end these waves will diffract around any normal obstacle; but as the frequency rises, they behave more and more like light, traveling in straight lines and clearly reflecting and refracting as light does.

Even at the highest radio frequencies that we can generate, the wavelength of electromagnetic radiation is many times longer than the longest wavelength of visible light. In between lies the region of heat radiation. This radiation is emitted by colliding molecules in hot gases. The thermal motions of the molecules involve accelerations of their charges with resulting emission of magnetic radiation of higher frequencies than we can produce with electrical circuits of man-made sizes.

At still higher temperatures—in arcs and sparks, for instance—the disturbed atoms emit light—infrared, visible, or even ultraviolet light. The visible part of the spectrum corresponds to frequencies of between 4×10^{14} cycles per second and 8×10^{14} cycles per second. Here we have long passed the region where we can directly measure the frequency, but we can identify the frequencies in light of a single spectral color by measuring the wavelengths. In this way, by using spectroscopes with the air pumped out, we can find the frequencies of the ultraviolet radiations emitted by atoms. These frequencies run up to a few times 10^{15} cycles per second.

For microwaves, infrared, visible light, and perhaps even ultraviolet, special intense phase-locked sources can be made. These are called masers (with an "m" for microwaves), irasers, or lasers. Their special nature was described briefly in Section 9–3. They have been made only with rather small total power so far, but that power is so concentrated in frequency and direction that the radiations have quite impressive effects on any absorbing matter.

Besides emitting light, atoms absorb it. An atom of a particular kind will absorb light of only certain very definite frequencies. Presumably, these are the frequencies of the natural motions of the electrons in the atom. The atom is tuned only to these frequencies and will respond only to electric fields which oscillate with them. We cannot tune atoms in the way that we can tune an oscillatory circuit that we build ourselves. But by looking for the right kind of atom, we may be able to find an electronic motion tuned in a frequency region in which we want to absorb light.

At higher frequencies we come to the region of X rays. Here we can make a crude estimate of the frequencies expected by calculating the time during which the high-speed electrons in an X-ray tube are decelerated when they crash into the target of heavy metal in the tube. We expect that the frequencies of the X rays generated will run up to approximately one divided by this time. If the frequency of the X ray is not too high, its wavelength will be about the size of an atom. Then we can use the regular gratings ruled in crystals by the natural repetitious spacing of the atoms to get a measurement of the wavelength. Also, indirect measurements, measurements that you will be able to understand from the discussion in Chapter 34, agree with the frequencies we compute from the crystal-grating measurements of wavelength. These indirect measurements can be extended to still higher frequencies, and they show that the X-ray frequencies observed are those we predict from our estimates of the stopping time of the accelerated charges.

Finally, at enormously high frequencies, we come to gamma rays. They are emitted spontaneously in some radioactive decay processes. Like light, they travel in straight lines at 3×10^8 m/sec and cannot be deflected by electric or magnetic fields. The same kind of measurement that we use to determine high X-ray frequencies shows that there are gamma rays with a frequency of 10^{21} cycles per second.

All these electromagnetic radiations carry energy. When they are absorbed, the body absorbing them heats up. This means that electromagnetic radiation, as we expected, can only be emitted in a process in which energy is supplied. As we assumed at the beginning, the radiating electric and magnetic fields must be set up when a force accelerates a charge. The evidence that electromagnetic radiation arises from accelerated charges is now overwhelming. Electromagnetic radiation exists; it is propagated with the speed of light, and it arises from accelerated charges.

QUESTIONS

1.* Where will the charges shown in Fig. 29 – 1 be located if the wire is moved in the opposite direction?

2.* (a) In which direction is the force (due to motion through the magnetic field) on charges in the side arms of the loop of Fig. 29 – 2?
(b) Does this force help push the charges around the loop?

3. (a) Analyze the situation of Fig. 29 – 5 entirely from the point of view of the forces on charges in the wire as it moves through the magnetic field, and show why there is no induced current.
(b) Continue this line of argument to show why there is induced current as one side of the loop gets into the weaker field at the edge of the magnet.

4.* In Fig. 29 – 5, would you expect induced current if the magnet were moved while the loop was held stationary?

5.* In Fig. 29 – 11, as the loop of wire rotates, at what orientation (a) is the maximum current induced, (b) does the induced current reverse direction?

6.* Why can't you induce a current in a loop of wire by moving it parallel to itself in a uniform field?

7.* A loop of wire with an area of 0.5 m^2 is in a uniform field of 2×10^{-2} newton/amp-m. What is the flux through the loop if the plane of the loop is perpendicular to the field?

8. What would be the magnetic flux inside a long solenoid 4.0 cm in diameter with 10 turns per cm of length when the current is 1.0 amp?

9.* If you were pulling the loop in Fig. 29 – 2 at a fixed speed, would you find it easier to pull it if its electrical resistance were increased?

10.* Does it take more work to pull a coil out of a magnetic field at high speed or at low speed?

11.* What EMF is induced in a 5-cm-long wire moving with a speed of 10 cm/sec across a field of 0.02 newton/amp-m?

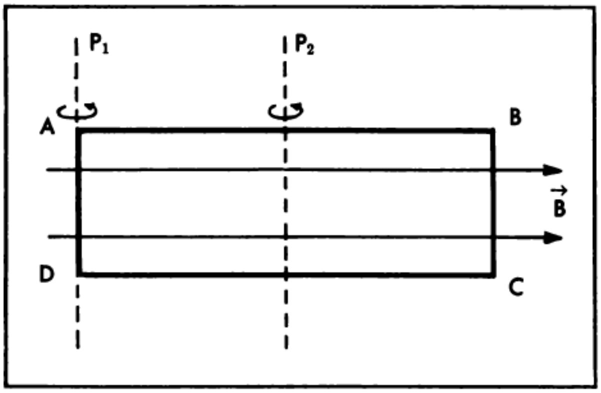
12.* What EMF is induced in a loop of wire if the flux through the loop changes at a rate of 0.5 newton-m/amp-sec?

13. Suppose that in Fig. 29 – 2 the magnetic field strength is 0.5 newton/amp-m, l is 10 cm, the resistance of the loop is 0.5 ohm, and the loop is moved at a constant speed of 2 m/sec.
    (a) What is the EMF generated in the wire?
    (b) What is the current in the circuit?
    (c) What is the magnetic force acting on the wire?

14. A metal helicopter rotor whose length from center to tip is 5.0 meters is rotating horizontally at 8.0 revolutions per second at a place where the earth's magnetic field has a vertical component downwards of 5.6×10^{-5} newtons per ampere-meter.
    (a) What is the induced EMF between the center and the tips?
    (b) Should this be of concern to the pilot?

15. Figure 28 – 17 is a diagram of an electric motor.
    (a) Explain how you could use the same arrangement of magnetic field and wire loop to make a generator that would produce current in only one direction in an external circuit.
    (b) How should the wire loop be oriented in the magnetic field when the commutator interchanges the connections between the ends of the loop and the wires of the outside circuit?

16. The rectangular wire loop shown in Fig. 29 – 22 is a uniform magnetic field.
    (a) In what direction (clockwise or counterclockwise) is the induced current when the loop is in the plane of the paper and *BC* is moving into the page about P_1 as an axis? About P_2?
    (b) For a given speed of rotation, how would the induced currents compare when the loop is rotated about P_1 and then about P_2?
    (c) For a given speed of rotation, how does the induced EMF depend on the area of the loop?
    (d) If the magnetic field has a magnitude of 1.5 newtons/amp-m, $AB = 10$ cm, and $BC = 4$ cm, what is the maximum induced EMF (in volts) when the loop is rotating about P_1 at 19.1 revolutions per second $\left(\frac{120}{2\pi}\text{ rps}\right)$?

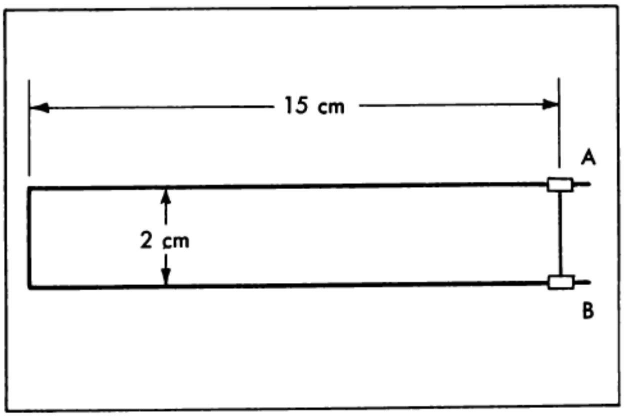


29–22 For Question 16.

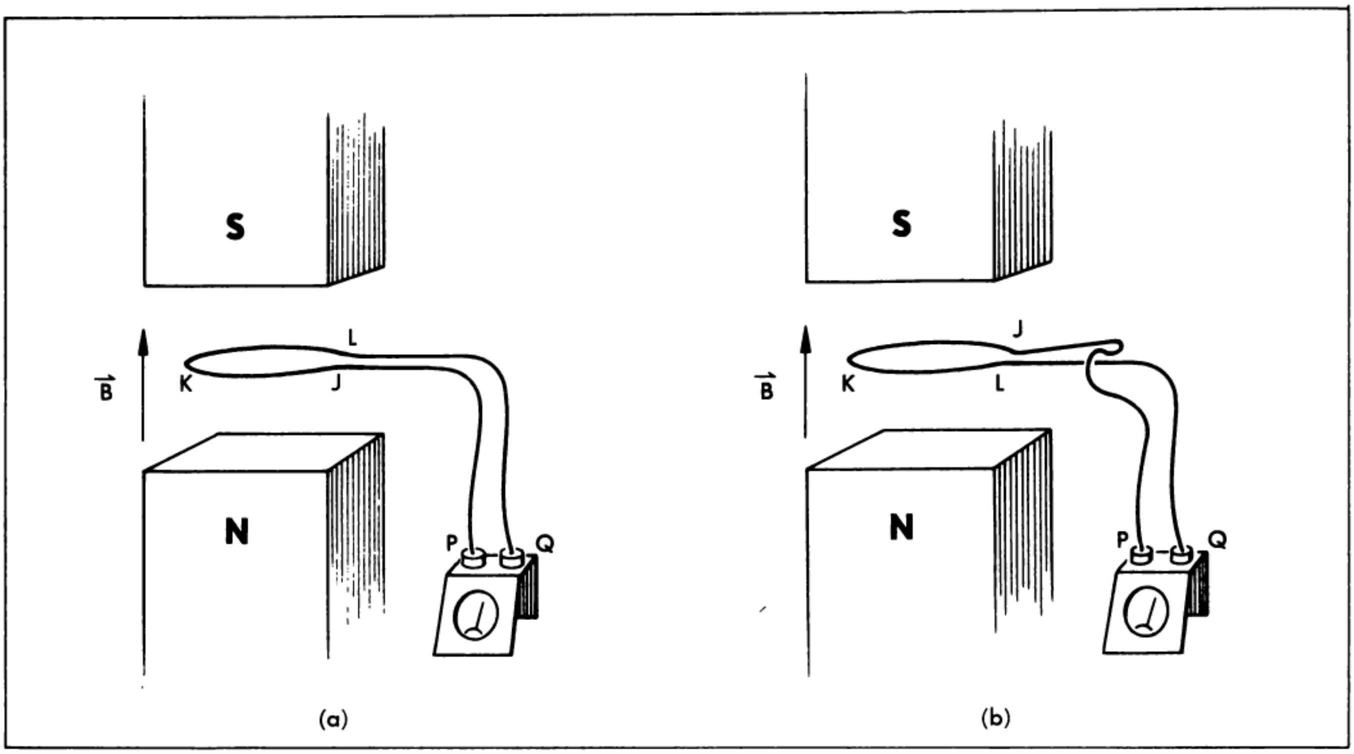
17. A U-shaped wire (Fig. 29 – 23) has a movable wire *AB* connected to it. This arrangement is in a uniform field perpendicular to, and into, the page.
    (a) If the magnetic field strength is 40 newtons/amp-m, what is the induced EMF (in volts) when *AB* is in the position shown and moving at 20 cm/sec? Calculate this first from the rate of flux change and then from the magnetic force on the elementary charges in the wire.
    (b) What is the induced EMF when *AB* is 5.0 cm from the left end and moving at 20 cm/sec? At 10 cm/sec?
    (c) At what rate is energy fed into the loop when *AB* moves at 20 cm/sec and the induced current is 2.0 amperes?
    (d) If the induced current is 2.0 amperes, what force is needed to move *AB* at 20 cm/sec?

18. Suppose that the moving circuit of Question 13 has a mass of 200 grams, and that the external force is suddenly removed when it is moving at 2 m/sec. What will its motion be like afterwards?

19.* If we applied a constant force to the loop in Fig. 29 – 2, would the loop accelerate uniformly as long as the segment of length *l* was in the magnetic field?



29–23 For Question 17.



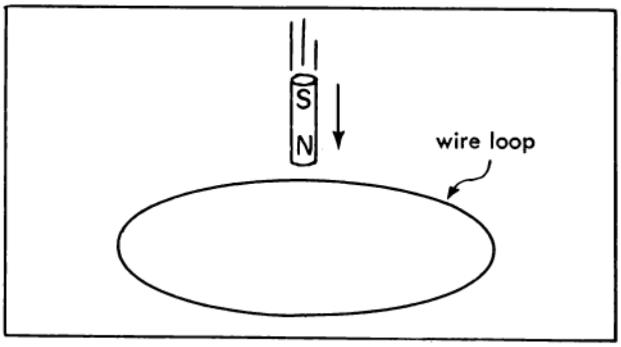
29–24 For Question 20.

20. A loop of wire *JKL* lies between the pole pieces of a magnet with the plane of the loop parallel to the pole pieces, as shown in Fig. 29 – 24 (a). It is connected to a sensitive ammeter whose pointer stands at the center of the scale when there is no current, and which is so constructed that the pointer swings toward the terminal (*P* or *Q*) at which the current enters the ammeter.

    (a) Which way does the pointer swing when the loop is pulled out of the field of the magnet? When it is put back as it is in Fig. 29 – 24 (a)?

    (b) Which way does the pointer swing when the loop is put back as it is in Fig. 29 – 24 (b)?

    (c) How does the pointer move when the loop is turned over from the first position to the second without being taken out of the field of the magnet?



29–25 For Question 22.

21. (a) Will the earth's magnetic field induce currents in an artificial satellite with a metal surface that is in orbit around the equator? Around the poles?

    (b) If so, how would these currents affect the motion of the satellite?

22. A bar magnet is dropped through a wire loop as shown in Fig. 29 – 25.

    (a) Describe the changes in the direction and magnitude of the induced current in the loop as the magnet falls through the loop.

    (b) Neglecting air resistance, will the acceleration of the falling magnet be constant?

23. A rectangular loop of copper wire is dropped between the poles of a magnet as shown in Fig. 29 – 26. Describe qualitatively the motion of the loop. Neglect air friction.

24.* Do the lines of an electric field induced by a change in magnetic flux have a beginning and an end?

25. In a betatron, electrons are accelerated to very high energies by a changing magnetic flux. The electrons move in a circle around the inside of an evacuated glass "doughnut" and speed up when the flux through the hole in the doughnut increases. A separate magnetic field passes through the doughnut perpendicular to its plane to keep the electrons moving in a circle.

An electron is traveling in a circle of 50 cm radius at 98 percent of the speed of light in the doughnut of a betatron. If the flux inside the doughnut increases at a rate of 24 newton-meters/amp-sec for $\frac{1}{240}$ second, how much energy (in electron volts) will be transferred to the electron? (*Note:* Assume that the speed stays constant at approximately the speed of light.)

26.* What determines the "equivalent current" between two parallel plates in an electric circuit?

27.* Is it possible to maintain indefinitely (a) a constant magnetic circulation? (b) a constant electric circulation?

28.* In Fig. 29 – 18 (c), assume that the plates started with zero charge and that the current shown has been flowing for a while. (a) What is the sign of the charges on each plate? (b) What is the direction of the electric field?

29. Suppose that at the moment there is no electric field where you are, but a short time later there will be an electric field pointing vertically down. The boundary is moving straight toward you. When it has reached you, what will be the direction of the magnetic field at your position?

30.* What is the speed within a vacuum of (a) radio waves, (b) infrared waves, (c) gamma rays?

31. A radio transmitter is producing electric oscillations with a frequency of 1.0×10^9 cycles/sec. The transmitter is connected to a small "dipole" antenna, as shown in Fig. 29 – 27. With the antenna in this horizontal position, the electric field lines are horizontal and the radiation is said to be horizontally polarized.

Suppose that you hold up a grid of copper rods supported by a plastic frame as shown in the figure. Will the radiation pass through the grid when the rods are horizontal? When they are vertical? Explain.

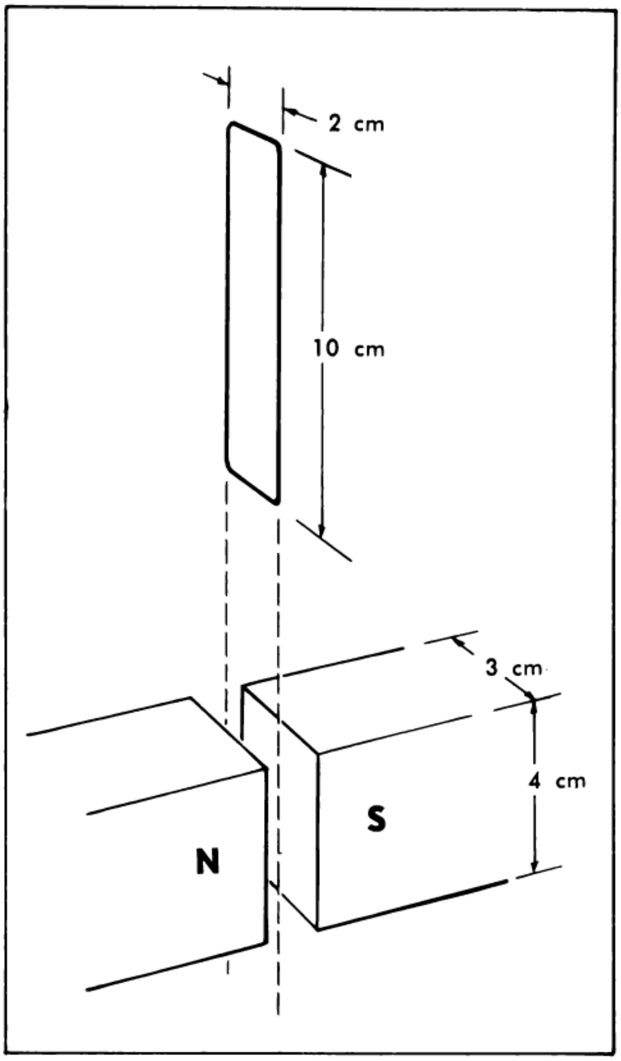
32. When electrons moving at 3×10^7 m/sec hit the target in an X-ray tube, they are brought to rest in about an atomic diameter – in about 10^{-10} m.

(a) What is the stopping time?

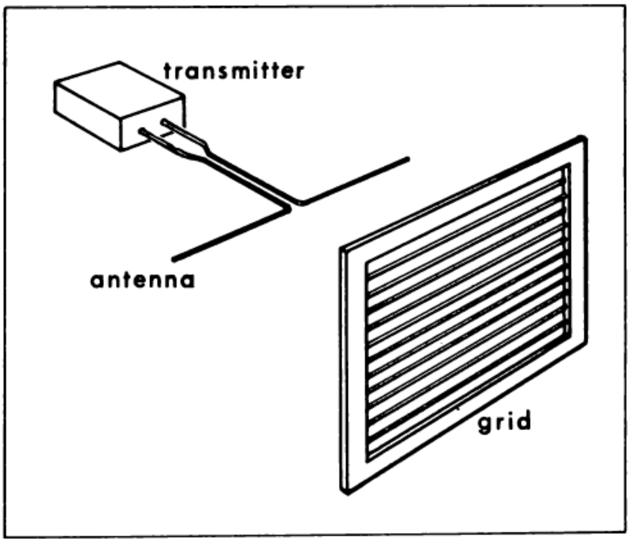
(b) If you use one divided by this time to estimate the frequencies you might expect to see in the electromagnetic radiation arising from this deceleration, what frequency do you get?

(c) What wavelength does this frequency give?

33. What evidence is there that sound is not electromagnetic radiation?



29–26 For Question 23.



29–27 For Question 31.

30 Relativistic Kinematics

30–1 A MEDIUM FOR LIGHT WAVES?

IN THE LAST section of Chapter 9 we raised several questions about waves in general and light in particular, but did not answer them. Let us now go back to one of them, that of the medium which carries the light waves, and investigate it in some detail.

The waves on a coil spring, water waves, and sound waves all propagate in a medium. There is always something, the shape of which moves. It is natural to ask what is the medium in which light waves travel. In the case of light waves it is the shape of the electric and magnetic fields that changes. Can the changes in these fields propagate through an absolute vacuum, or do they also require a medium?

This question puzzled many physicists of the nineteenth century, and they devised various experiments in order to prove the presence of the light-carrying medium—the "ether." They realized that the ether must be very different from all other wave-carrying media because it apparently is present even in the highest vacuum as well as in transparent materials. It is therefore unlikely that the ether is a form of matter with such properties as chemical composition, density, and the like. The physicists of the nineteenth century did not look for these material properties but asked themselves the following fundamental question:

The ether occupies all space out to the farthest stars. The earth moves through this space, rotating around its axis and around the sun. How does the ether move with respect to the earth? Does the ether follow the earth's motion, being therefore at rest with respect to the earth, or is the ether at rest with respect to the sun and the other fixed stars? In the latter case, it is obvious that the ether would move with respect to the earth.

It certainly is presumptuous for us to believe that the ether which fills the whole universe follows the motion of one particular planet. There also is experimental evidence which contradicts the assumption that the ether is at rest with respect to the earth. What, then, about the possibility that the ether is at rest with respect to the fixed stars? Can we detect its motion relative to the earth?

30–2 WAVES IN MOVING MEDIA

Imagine that you stand on the bank of a river which flows by very evenly. How can you find out whether the water is really flowing by—and, if it is, measure its speed? Try to think of a method that can also be used for detecting the motion of the ether relative to the earth. Throwing a cork in the water to see if the water drags it along, or putting a finger in the water to feel a push in the direction of motion of the water, will not do. We cannot do similar experiments with the ether: it would not drag objects along and would not exert a push on them. We must restrict ourselves to experiments involving only the waves.

Suppose we are equipped with a wave generator, meter stick, clock, and stroboscope. One experiment comes readily to mind. We generate a straight pulse at a point A along the bank; measure the time it takes to get to point B, a distance l away (Fig. 30–1); and compare it with the time it takes a similar pulse to go from B to A. If the water in the river is not flowing at all, and the pulse propagates with speed u, then we expect both times to be equal to l/u.

If the river flows, say, to the right, a pulse will require less time to go from A to B than from B to A. Relative to an observer who moves along with the water, the pulse moves with velocity u. But if the velocity of the water itself relative to the bank is v, then the velocity (relative to the bank) of the pulse moving from A to B is $u + v$, and hence the time required to cover the distance is

$$t_{AB} = \frac{l}{u+v} \quad \text{or} \quad u + v = \frac{l}{t_{AB}}$$

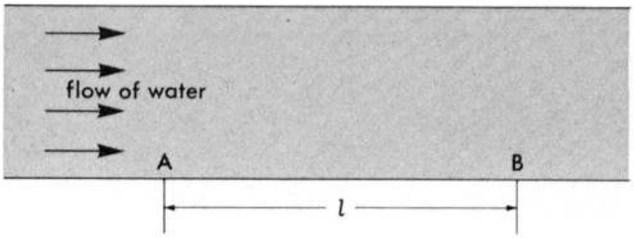
Similarly, the time needed for a pulse from B to get to A is

$$t_{AB} = \frac{l}{u-v} \quad \text{or} \quad u - v = \frac{l}{t_{BA}}$$

Subtracting the second equation on the right from the first one, we find

$$2v = \frac{l}{t_{AB}} - \frac{l}{t_{BA}} = \frac{l}{t_{AB}t_{BA}}\,(t_{BA} - t_{AB})$$

To calculate v we must be able to measure t_{BA} and t_{AB} with sufficient accuracy that the experimental error will be less than the difference $t_{BA} - t_{AB}$. For example, if $t_{BA} - t_{AB}$ is 10 percent of, say, t_{BA}, and both t_{BA} and t_{AB} are measured to within 0.1 percent, then inserting the measured values into the last equation will yield v within 2 percent, a meaningful result. It is not difficult to measure the velocity of the water in a river by this method, because the velocity of the water, v, is often a fair fraction of the wave velocity, u, and therefore t_{BA} and t_{AB} are markedly different.



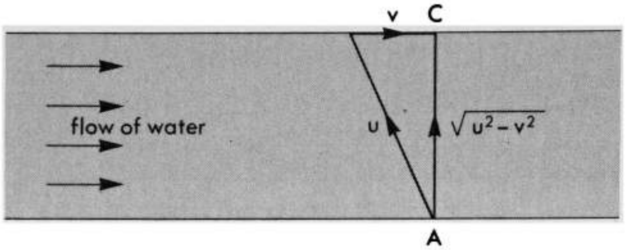
30–1 With water flowing to the right, it takes less time for a pulse to go from A to B than from B to A.

The situation is quite different when we try to perform the analogous experiment with light.

30–3 A ONE-WAY MEASUREMENT OF THE SPEED OF LIGHT

The orbit of the earth around the sun is nearly a circle of circumference 9.36×10^{11} m, and it takes the earth one year or 3.16×10^{7} sec to complete one revolution. The mean speed of the earth in its orbit is therefore 3.0×10^{4} m/sec. With the speed of light $c = 3.0 \times 10^{8}$ m/sec, the speed of the earth is only one part in 10^4 of the speed of light. If we let a pulse travel from a point A to a point B in the direction of the motion of the ether, we expect its speed to be greater than in the opposite direction by about two parts in 10^4. To detect this small difference we must be able to measure the travel time of the light pulse to better than one part in 10^4—say, one in 10^5.

This calls for two synchronized clocks running at the same rate to a very high but attainable accuracy. If two identical clocks are placed right next to each other, we can establish that they are synchronized by simply looking at the hands: if we see that the two sets of hands always read the same, we can say that the two clocks are synchronized. We could then



30–2 For the pulse from A to reach C, it has to be aimed somewhat upstream.

move one of the clocks as far away as we wish and use the clocks to measure the time it takes a light signal to travel in each direction between them. In so doing, we assume that moving a clock from one place to another does not affect its behavior.

We might try to check the validity of this assumption by synchronizing the two clocks in a different way after they have been placed in their proper locations. We place a flash bulb halfway between the two clocks and fire it. Each of the clocks starts running when the flash of light reaches it. Of course, this procedure will synchronize the clocks only if the speed of light is the same in both directions. But this is precisely what we want to check with the clocks, so we cannot use this method to synchronize them or to check whether moving a clock affects its behavior.

Fortunately, we do not need two clocks. We are really not interested in the individual times of propagation of the light signal in the two directions, but only in their difference. Therefore we should look for an experiment which measures the difference directly.

30–4 THE MICHELSON-MORLEY EXPERIMENT

Let us go back for a moment to the river (Fig. 30 – 1). The time required to go from A to B and back is

$$t_{\parallel} = t_{\mathrm{AB}} + t_{\mathrm{BA}} = \frac{l}{u+v} + \frac{l}{u-v} = \frac{2lu}{u^2 - v^2}$$

We can rewrite this expression in terms of the time $2l/u$ it would take to make the round trip if the water were not moving:

$$t_{\parallel} = \frac{2l}{u} \cdot \frac{1}{1 - v^2/u^2}$$

Now, assume that we want to send a straight pulse in the direction perpendicular to the flow, say to a point C opposite to A on the other side of the river (Fig. 30 – 2). If we send the pulse toward C it will not get there, because it is shifted by the motion of the water. We have to aim the pulse somewhat upstream so that by the time it reaches the other side it has just been shifted to C (Fig. 2). Since the pulse moves with the speed u with respect to the water, it has the speed $u' = \sqrt{u^2 - v^2}$ with respect to the observer on the bank. At C the angle of reflection equals the angle of incidence. The pulse again propagates somewhat upstream, while the water shifts it downstream, and it returns to A with the same speed relative to the bank, $u' = \sqrt{u^2 - v^2}$. The round trip will, therefore, take a time:

$$t_{\perp} = \frac{2l}{\sqrt{u^2 - v^2}} \quad \text{or} \quad t_{\perp} = \frac{2l}{u} \cdot \frac{1}{\sqrt{1 - v^2/u^2}}$$

The difference between the two times:

$$t_{\parallel} - t_{\perp} = \frac{2l}{u} \cdot \frac{1}{\sqrt{1 - v^2/u^2}} \left(\frac{1}{\sqrt{1 - v^2/u^2}} - 1 \right)$$

can be measured directly.

For the motion of the earth through the ether we can simplify the expression for $t_{\parallel} - t_{\perp}$. Here the speed v of the earth in its orbit around the sun is much smaller than the speed of the waves, which is the speed of light, c. (This is commonly written as $v \ll c$ or $v/c \ll 1$.) Hence $1/\sqrt{1 - v^2/c^2}$ is very close to 1; and the expression inside the parentheses can be approximated by $v^2/2c^2$. (See box for derivation.) Thus,

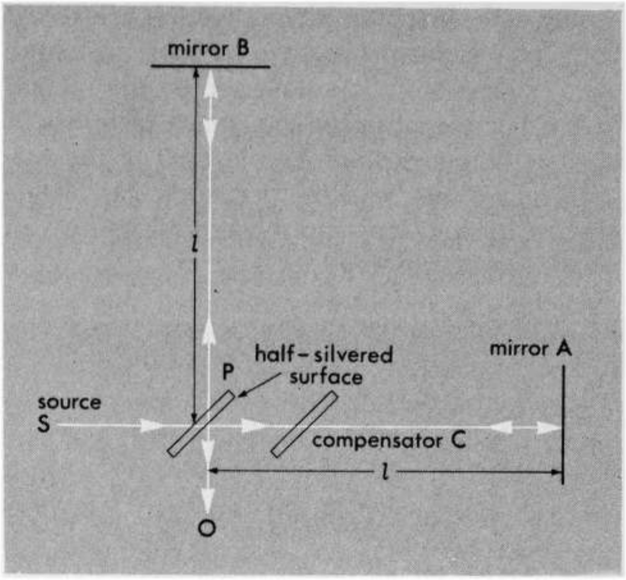
$$t_{\parallel} - t_{\perp} = \frac{l}{c} \cdot \frac{v^2}{c^2}$$

In 1887 Michelson and Morley performed an experiment to measure $(t_{\parallel} - t_{\perp})$ for light. They compared the time it takes a light signal to go back and forth along the line of motion of the ether relative to the earth with the time it takes it to complete the round trip in the perpendicular direction. In their experiment light traveled a distance of $l = 11$ meters in each direction. As we saw in Section 3, $v/c = 1.0 \times 10^{-4}$. Hence the expected time difference was

$$t_{\parallel} - t_{\perp} = \left(\frac{11 \text{ m}}{3.00 \times 10^8 \text{ m/sec}} \right) (1.0 \times 10^{-4})^2$$
$$= 3.7 \times 10^{-16} \text{ sec}$$

This is an extremely short time. How could they measure it? To measure short times requires a clock with a short period. Michelson and Morley used the period of monochromatic light waves as the period of their clock. Fig. 30 – 3 shows a schematic diagram of their apparatus.

Light from the source *S* hits a glass plate (*P*) set at 45°. The rear surface of the glass plate is half-silvered so that approximately half of the light is reflected and half is transmitted. The transmitted light travels a distance $l = 11$ m to mirror *A* and is reflected back to the half-silvered surface. There, part of it is reflected into the eyepiece *O*. The light originally reflected by the half-silvered surface also travels a distance of 11 m, is reflected by mirror *B*, and part of it is transmitted by the half-silvered surface into the eyepiece. A second glass plate (*C*)



30–3 A schematic drawing of the interferometer used by Michelson and Morley.

Power Expansions

In physics we often deal with mathematical expressions like $\frac{1}{1-x}$, $\frac{1}{1-x^2}$, $\frac{1}{\sqrt{1-x^2}}$, etc., where x is a pure number much smaller than 1. In this chapter, for example, we have $x = \frac{v}{c} \ll 1$. Approximations to these expressions are useful.

The expression $\frac{1}{1-x}$ can be expanded into a power series in x by repeatedly using the identity $1 = 1 - x + x$ in the following way:

$$\begin{aligned}\frac{1}{1-x} &= \frac{1-x+x}{1-x} = 1 + x\ \frac{1}{1-x}\\ &= 1 + x\ \frac{1-x+x}{1-x} = 1 + x + x^2\ \frac{1}{1-x}\\ &= 1 + x + x^2 + x^3\ \frac{1}{1-x}\end{aligned}$$

This expansion can be carried on to any number of terms. As you would expect, it gives the same result as does ordinary long division. For $x \ll 1$ each term is significantly smaller than the one preceding it.

Notice that each term contains a higher power of x than the one preceding it. We generally speak of the term not containing x as the "zero order term," the term containing x to the first power as the "first order term," etc.

When x is a physical quantity it is meaningless to calculate $\frac{1}{1-x}$ to a higher order than the accuracy of x justifies. It often suffices to retain only the first order term:

$$\frac{1}{1-x} \approx 1 + x$$

Similarly, $\frac{1}{1-x^2} \approx 1 + x^2$ to first order in x^2, or second order in x. There is no first order term in x.

Another useful relation which holds to first order is $\sqrt{1-x} \approx 1 - \frac{x}{2}$. To see that this is true, we square the right-hand side and find $\left(1 - \frac{x}{2}\right)^2 = 1 - x + \frac{x^2}{4}$, which differs from $1 - x$ only in second order.

It follows from the above arguments that

$$\frac{1}{\sqrt{1-x^2}} \approx \frac{1}{1 - \frac{x^2}{2}} \approx 1 + \frac{x^2}{2}$$

to second order in x. With $x = \frac{v}{c}$ we have

$$\frac{1}{\sqrt{1-v^2/c^2}} - 1 \approx \left(1 + \frac{1}{2}\,\frac{v^2}{c^2}\right) - 1 = \frac{1}{2} \cdot \frac{v^2}{c^2}$$

similar to the first but with no silvering, is inserted in the path of the first beam to compensate for the extra thickness of glass traversed by the second beam. For the observer the image of mirror *A* appears at the position of mirror *B*. If the two mirrors are not exactly perpendicular, interference fringes similar to those produced by an air wedge are observed.

Michelson and Morley mounted this apparatus on a smooth bearing and pointed the arm holding mirror *A* in the direction of motion of the earth around the sun, that is, in the direction opposite to the flow of the ether. The arm carrying mirror *B* was perpendicular to the ether flow. If the ether is indeed at rest with respect to the sun, then the light reflected from mirror *A* will require a longer time to reach the eyepiece than that from mirror *B*. Therefore, the images of the light source *S* in the two mirrors will act like two sources with a fixed phase delay between them. The phase delay will be

$$p = \frac{t_{\parallel} - t_{\perp}}{T}$$

where T is the period of the light. For the yellow line of sodium, $T = 20 \times 10^{-16}$ sec; hence, $p = 0.19$.

Now, when the apparatus is rotated through an angle of 45°, the travel time of the two light beams will be the same and the images of *S* will appear as two light sources in phase. Finally, when the apparatus is turned through 90°, there is again a phase shift of 0.19 but this time the light from *B* will lag behind that from *A*.

We know from the study of interference (see Section 8 – 5) that a change in the phase of the sources causes a shift in the interference pattern. Specifically, a change of phase of unity shifts all lines to the position of the adjacent lines. Michelson and Morley argued that when they turned their interferometer through 90° the interference fringes should shift a fraction $2 \times 0.19 = 0.38$ of the distance between consecutive fringes.

Such a shift can hardly escape detection; nevertheless, it was not there. To be sure, there were minute shifts of the interference pattern as the interferometer was rotated, but those were not related in any consistent way with the orientation of the arms, and could be accounted for by small variations in the temperature and by mechanical deformations.

The Michelson-Morley experiment has been repeated many times. But the predicted effect was never observed.

30–5 THE LORENTZ CONTRACTION

It was clear to physicists at the turn of the century that a bold new idea was required to account for the null result of the Michelson-Morley experiment. The Dutch physicist H. A. Lorentz presented the following argument: According to the calculation, the beam traveling in a direction perpendicular to the ether flow requires a time

$$t_{\perp} = \frac{2l}{c} \cdot \frac{1}{\sqrt{1 - v^2/c^2}}$$

whereas the time for the beam parallel to the flow is

$$t_{\parallel} = \frac{2l}{c} \cdot \frac{1}{1 - v^2/c^2}$$

If the arm of the interferometer parallel to the ether flow were shortened by a factor $\sqrt{1 - v^2/c^2}$, while the arm perpendicular to the ether flow remained unchanged, then $t_{\parallel}$ would be reduced by $\sqrt{1 - v^2/c^2}$ and would become

$$\begin{aligned} t_{\parallel} &= \frac{2l\sqrt{1 - v^2/c^2}}{c} \cdot \frac{1}{1 - v^2/c^2} \\ &= \frac{2l}{c} \cdot \frac{1}{\sqrt{1 - v^2/c^2}} = t_{\perp} \end{aligned}$$

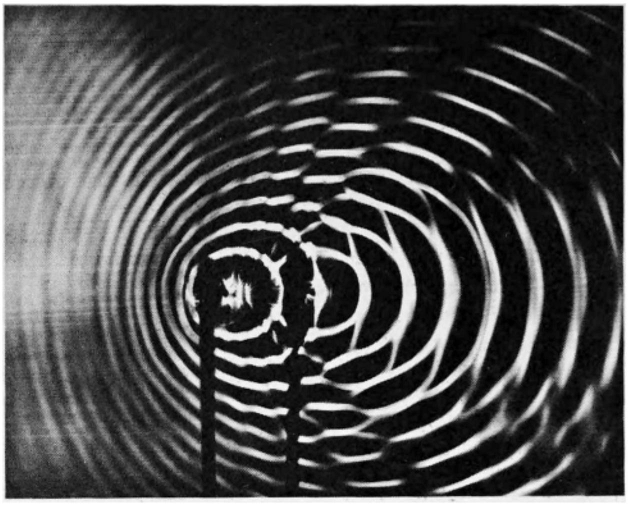
Lorentz suggested that an object moving through the ether with speed v is actually contracted by a factor $\sqrt{1 - v^2/c^2}$ in the direction of motion, while its linear dimensions in the plane perpendicular to the motion remain unchanged. This contraction cancels the effect caused by the dependence of the speed of light on the direction and thus accounts for the null result of the Michelson-Morley experiment.

As bold as this idea must have appeared at the time, it still seemed less revolutionary than dropping the notion of the ether. We shall see later in this chapter that the Lorentz contraction follows from a more general approach to the question of the propagation of light. Nevertheless, as a basic assumption, it is not sufficient; while it explains the Michelson-Morley experiment, it fails to explain others. In the next section we shall discuss such an experiment involving the effect of the earth's motion on a double-slit interference pattern.

30–6 TWIN-SOURCE INTERFERENCE IN A MOVING MEDIUM

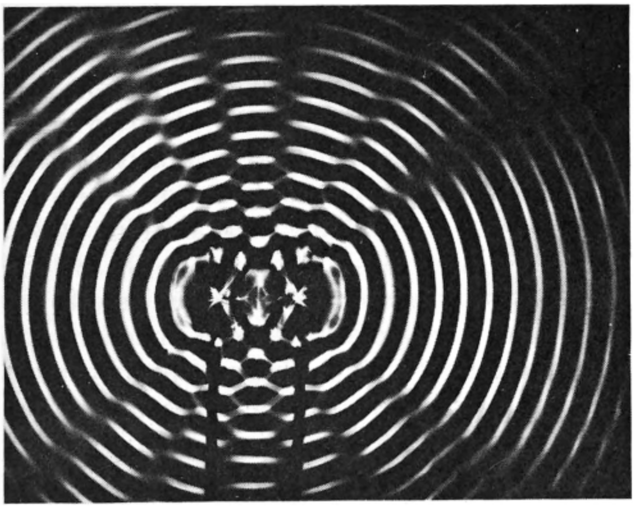
To clarify the effect of the earth's motion through the ether on a double-slit interference pattern, let us study first a similar situation in the ripple tank. Figure 30 – 4 shows the familiar interference pattern produced by the two point sources in phase. Figure 30 – 5 shows the pattern produced by the same sources when the water moves to the right relative to the sources. Notice how the whole pattern is distorted and bends toward the direction of the flow of the water. In this picture the speed of the water, v, was a large fraction of the speed of the waves, u; in fact, $v/u = 0.7$. Keeping in mind that we want to apply the results of the present study to the earth's motions through the ether, we can confine ourselves to much smaller values of v/u.

Figure 30 – 6 shows the interference pattern for $v/u = 0.2$. Here there is hardly any distortion; there is just a mild shift of the pattern as a whole, similar to the shift produced by a phase delay between the sources. (See Fig. 8 – 17.) To

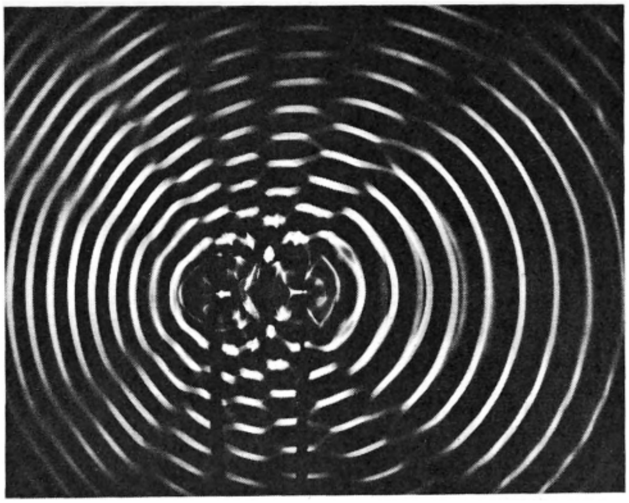


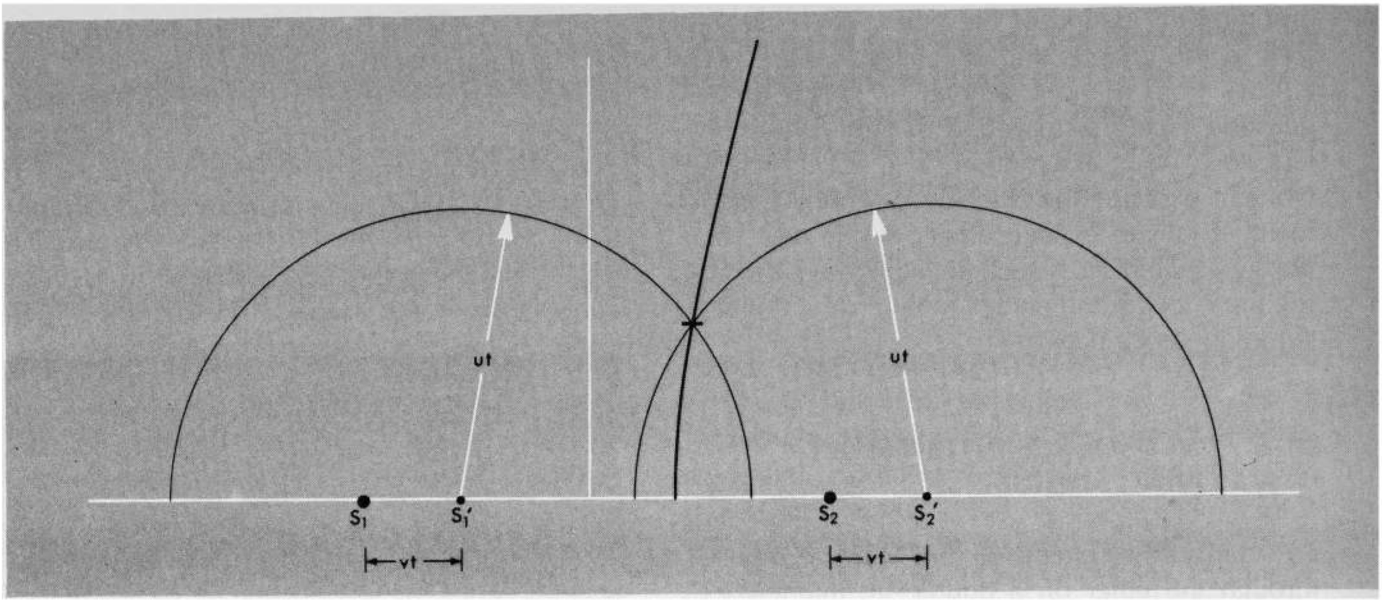
30–5 The same sources as in Fig. 30–4, but with the water moving to the right at a speed which is a large fraction of the speed of the waves themselves.

30–4 The interference pattern of two sources in phase.

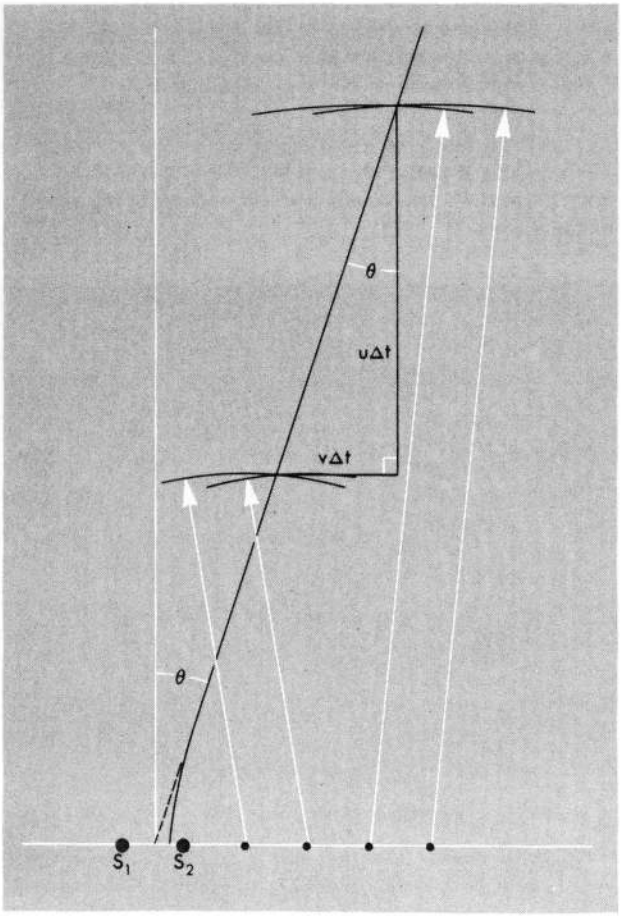


30–6 The same sources as in Fig. 30–4 and 30–5 but with the water's speed only a small fraction of the speed of the waves.





30–7 Circular pulses emitted simultaneously by S_1 and S_2 are carried to the right by the flowing water. A time t after emission, the centers of the crests have moved to S_1' and S_2', respectively. The crests intersect to form a double crest to the right of the perpendicular bisector of S_1S_2. The heavy curved line shows the path of the double crest.



30–8 The intersection of the same two crests shown at two instants, Δt apart, long after the moment pictured in Fig. 30–7.

express this shift quantitatively, it will suffice to find the shift of the central maximum. The central maximum is mapped out by the point of intersection of expanding crests (or troughs) which are emitted simultaneously. Figure 30 – 7 shows the position of such a double crest at a time *t* after emission, as seen in the frame of reference in which the two sources are at rest. In this frame the water moves to the right with speed *v*. At the instant shown, the two crests are circles of radius *ut* whose centers are displaced downstream by a distance *vt*.

Figure 30 – 8 shows the same situation at two times long after those illustrated in Fig. 30 – 7. The two segments of the individual crests are now moving in almost the same direction, and the double crest path has become a straight line making an angle θ with the normal direction. Thus the effect of the moving medium is to shift

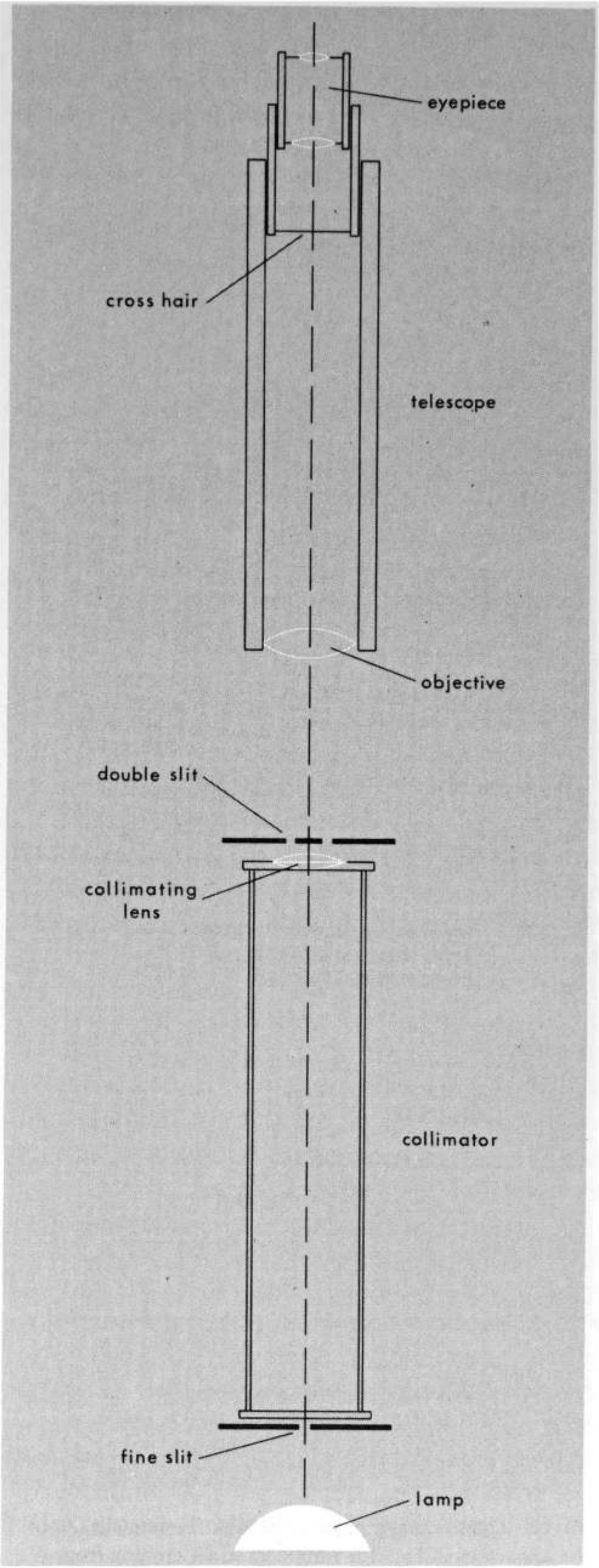
the central maximum of the interference pattern downstream by an angle θ. Clearly, $\tan \theta = v/u$; for small v/u, the angle θ is small too, and we simply write $\theta = v/u$, where θ is expressed in radians.

Now imagine the two sources are in the middle of a river that is flowing uniformly with speed v, in a direction parallel to the line connecting the two sources. We hold a glass plate over the water, make a reference line on it, and draw the interference pattern, marking especially the central maximum. If the water flows to the right, the central maximum is shifted to the right. Now we turn with our glass plate through 180°, so that the water now flows to our left with the same speed, and again mark the central maximum. The directions of the central maxima as marked on the glass will form an angle $2\theta = 2v/u$.

Thus, in this experiment the motion of the medium causes a shift in the interference pattern which is proportional to v/u. This is a first-order effect.

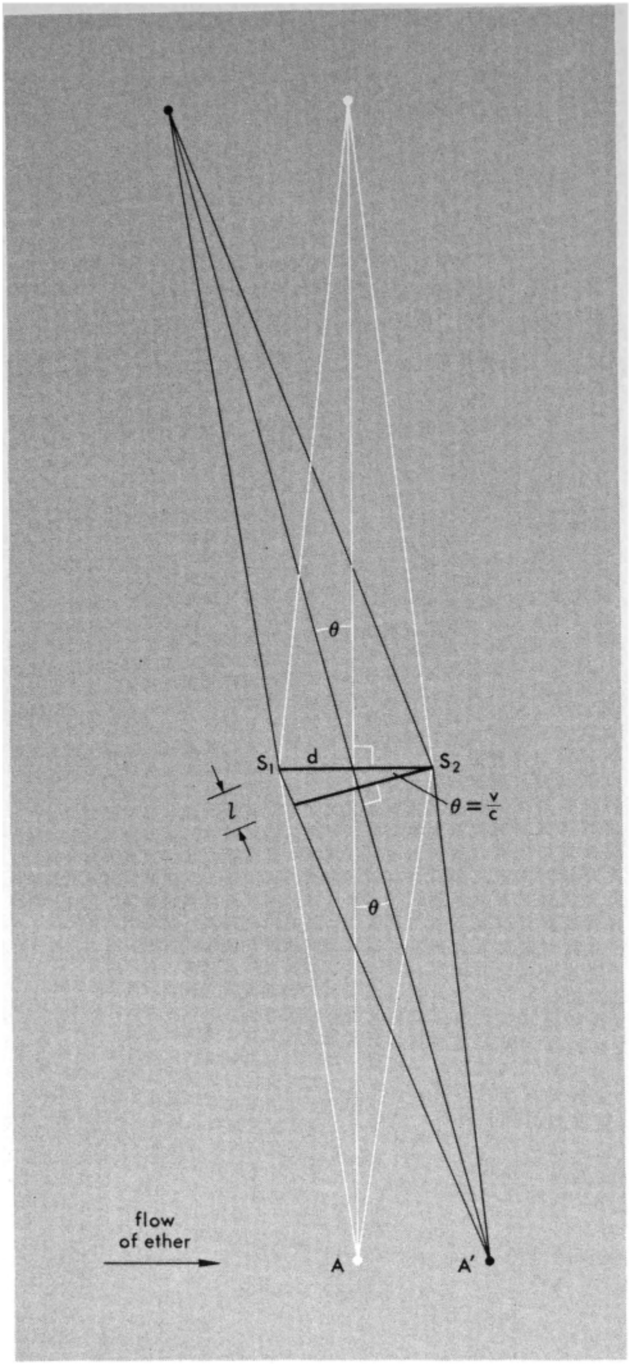
30–7 A FIRST-ORDER EXPERIMENT WITH LIGHT

If we want to perform an experiment with light waves which is exactly analogous to the experiment we have just described, we run into a difficulty: we do not have two light sources which will stay in phase. (See Section 9 – 3.) We have overcome this difficulty by illuminating two slits and using them as two light sources locked in phase. Thus, to observe the expected shift of a double-slit interference pattern we might use an instrument like the one illustrated in Fig. 30 – 9. The single slit is at the focus of the lens. The light which passes through the slit leaves the lens as a parallel beam and strikes the double slit. We observe the interference pattern through a telescope at the other end. A cross hair in the telescope takes the place of the reference line on the glass plate in the water experiment mentioned above. But with this arrangement the interfering light waves start from a single source and not from two independent sources, and they pass through the ether all the way. The consequences of this



30–9 Schematic drawing of apparatus to observe the shift of an interference pattern due to the earth's motion relative to the ether.

situation can be seen in Fig. 30 – 10. By the time a circular crest emitted at A reaches the slit S_2, the center of the circle will have moved to the point A' due to the motion of the ether. To arrive at S_1 the crest has to travel the additional distance



30–10 Light actually passing through the fine slit (A) of the apparatus in Fig. 30–9 appears to be coming from A', owing to the ether flow. It arrives at the double slit out of phase. If no ether were present on the other side, the central maximum would be shifted to the left. However, because ether is present on both sides of the double slit, there is no net effect.

$$l = A'S_1 - A'S_2 = d \sin \theta \approx d \frac{v}{c}$$

This requires an additional time

$$\Delta t = \frac{l}{c} = \frac{dv}{c^2}$$

The two slits therefore act as two sources out of phase (with S_1 lagging with respect to S_2) with a phase delay of

$$p = \frac{\Delta t}{T}$$

where $T = \lambda/c$ is the period of the light wave. Substituting the expressions for Δt and T, we have

$$p = \frac{dv}{c^2 T} = \frac{d}{\lambda} \cdot \frac{v}{c}$$

If there were no ether flow beyond the double slit, then this phase delay would cause a shift of the entire interference pattern to the left by an angle

$$\bar{\theta} = p \frac{\lambda}{d} = \frac{v}{c}$$

(See Chapter 8, Section 5.) Since we assume the same ether flow on both sides of the double slit, we expect a shift in the interference pattern to the right by an angle $\theta = v/c$, as we have seen in the preceding section. Thus the shift in the pattern to the left due to the difference in phase will just cancel the angular shift to the right due to the ether flow beyond the double slit. Hence there should be no effect; the interference pattern would be the same as if the medium were not moving at all.

We can expect an effect, however, if we insert a transparent material—water, for example—between the single slit and the double slit. In the water, the speed of light is c/n, where n is the index of refraction of the water. Hence, in water

$$\Delta t = \frac{dv}{c^2} \cdot n^2$$

and S_1 will lag behind S_2 by a phase delay of

$$p = \frac{\Delta t}{T} = \frac{d}{\lambda} \cdot \frac{v}{c} n^2$$

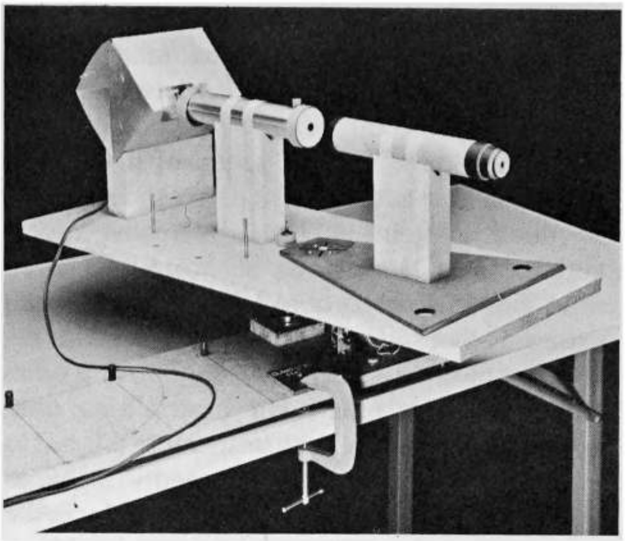
If these waves now pass into the ether at rest, the interference pattern will shift to the left by an angle vn^2/c. The motion of the ether to the right will reduce the shift by v/c just as before. Thus, the expected net shift is

$$\theta = \frac{v}{c}(n^2 - 1)$$

to the left. By rotating the apparatus through 180°, we double the effect:

$$2\theta = 2\frac{v}{c}(n^2 - 1)$$

It is quite simple to choose the wavelength of the light and the distance between the slits in such a way that the angle θ will correspond to a sizable fraction of the angle subtended by neighboring fringes and be easily detectable. You can do it yourself with the apparatus shown in Fig. 30 – 11. This experiment has been done, but no shift has been observed!



30–11 A modification of the apparatus shown in Fig. 30–9. The collimator tube between the light source and the double slit is filled with water. The whole apparatus can be rotated around a vertical axis.

30–8 THE SPEED OF LIGHT: A UNIVERSAL CONSTANT

Let us summarize what we have done in the last two sections. We discussed an interference experiment in the ripple tank which reveals the motion of the water with respect to the source and observer. We then designed a similar experiment with light in order to detect the earth's motion relative to the ether. Because of the difficulty of providing two independent light sources which will remain in phase, we used a more complicated arrangement in which we placed a refractive material in a section of the light's path. There was no effect with water as the refractive material. It is hard to believe that we would find an effect with, say, mineral oil. This strongly suggests that the refractive material is really not an essential part of our basic experiment, and that had we been able to produce two light sources in phase (without resorting to a single slit), we would also not find any effect. If there is an ether, we somehow failed again to detect our motion relative to it.

The null results of the Michelson-Morley experiment and of the first-order experiment described in the last section clearly demonstrate that the time it takes light to travel a given distance is the same for all directions. Hence, the speed of light in vacuum is also the same for all directions and has no relation to the earth's motion around the sun. (We say the speed of light is "isotropic" in the earth's frame of reference.) We started by assuming that this would be the case only in the sun's frame, that is, in the frame in which we thought the ether is at rest, and we ended by finding it also to be the case in the earth's frame of reference. There can hardly be anything special about the earth's frame of reference. Thus the speed of light is likely to be independent of direction in all frames of reference. We are led, therefore, to conclude that the ether is at rest in all frames, and this is hardly a helpful idea.

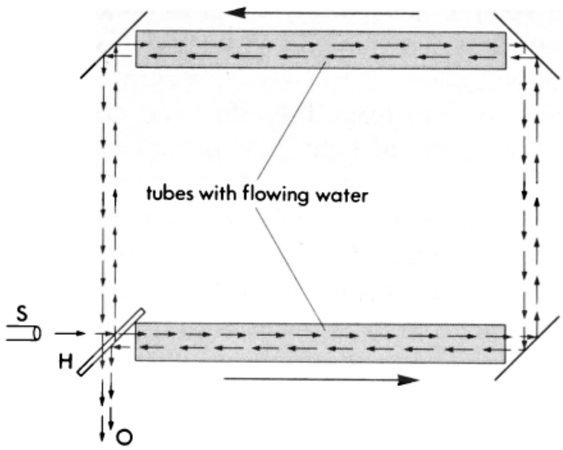
So we shall leave the ether and concentrate on the speed of light. We proved experimentally that the speed of light is isotropic in the earth's frame of reference and presented reasons for believing that the same holds in other frames which move relative to the earth. But what about the numerical value of the speed of light in vacuum? Is it the same in all frames of reference, or is it different, when measured by the same procedure, in different frames of reference? Is there, perhaps, a frame of reference in which the speed of light is largest? Naturally, we have no direct measurements of the speed of light in frames of reference other than our own.

But the evidence that the speed of light is uniform in all directions in various frames makes it plausible that the speed itself is also the same in all frames.

The conjecture that the speed of light is the same in all frames of reference, which was made by Albert Einstein in 1905, is really a very startling one. We know of no other speed which has this property. All other speeds, such as the speed of sound or of a rocket, are different in different frames of reference moving with respect to each other. For example, a train moving with speed v with respect to the tracks appears to be going with the speed $v + u$ with respect to another train going with speed u in the opposite direction. There is a great wealth of experience to support this simple addition of velocities. Nevertheless, this simple addition appears to fail when it comes to the speed of light.

30–9 THE FIZEAU EXPERIMENT

Think of the experimental evidence we have that does support the rule of simple addition of velocities. All the speeds are really quite small compared to the speed of light in vacuum. The speeds of the fastest bullets, for example, are of the order of 10^3 m/sec, less than 10^{-5} times the speed of light in vacuum. What happens to the rule when one of the speeds involved is close to the speed of light in vacuum? An answer to this question is found in an experiment performed over a hundred years ago by the French physicist Fizeau; it was later repeated independently by Michelson and Morley and by Zeeman.



30–12 The optics of Fizeau's apparatus. The plumbing that permits the rapid flow of water without blurring the interference pattern, though by no means unimportant, is not shown.

The speed close to that of light in vacuum was the speed of light in water; the other speed was the speed of the water itself with respect to the observer. A schematic drawing of the apparatus is shown in Fig. 30 – 12. Light from the source S hits the half-silvered mirror H. The transmitted part proceeds in a counterclockwise path until it hits the half-silvered mirror again; then half of it continues to the observer O. The reflected part proceeds clockwise until it is reflected again at H in the direction of the observer, where it interferes with the part which went counterclockwise. The resulting interference pattern is similar to that obtained with the Michelson interferometer. If we now place two water-filled tubes with glass ends in the path of the light beams, as shown in the figure, the interference pattern will not be affected; both the clockwise and the counterclockwise beams are equally delayed by the water. But what will happen to the interference pattern if we connect the pipes with each other and with a pump which will circulate the water?

If the water circulates counterclockwise, then the light waves traveling counterclockwise should get to the half-silvered mirror ahead of the waves traveling clockwise. This will cause a shift in the interference pattern. Let us calculate this shift, assuming that the simple addition of velocities, which we know holds for low speeds, holds also in this case.

Let the length of each tube be l and the speed of light in water be u. Then the time required for the light to pass through both tubes when the water is at rest is $2l/u$. If the water flows counterclockwise with speed v, then the speed of light going with the flow will increase to $u + v$ and that going against the flow will decrease to $u - v$. Thus the parts of a crest which was split by the half-silvered mirror will arrive again at the half-silvered mirror with the time delay

$$\Delta t = \frac{2l}{u-v} - \frac{2l}{u+v} = 4l\,\frac{v}{u^2 - v^2}$$

Since $u = c/n$, where n is the index of refraction of water, then $v \ll u$ and we can write for Δt:

$$\Delta t = \frac{4lv}{u^2} = \frac{4l}{c} \cdot \frac{v}{c} n^2$$

In Fizeau's original apparatus the tubes were 1.49 m long and the speed of the water was 7.1 m/sec. With $n = 1.33$ this gives $\Delta t = 0.83 \times 10^{-15}$ sec. The predominant color in his experiment corresponded to a period of $T = 1.75 \times 10^{-15}$ sec. Thus the phase delay between the two halves of the beam is $p = \Delta t/T = 0.47$.

Our calculation therefore predicts that, as the speed of the water is increased from zero to 7.1 m/sec, the interference pattern should shift by 0.47 of a fringe. Fizeau observed a shift of 0.23, only about half the expected value. This observation was confirmed by Michelson and Morley and by Zeeman. We see, therefore, that the light moving with the water flow is speeded up and the light moving against it is slowed down, but by an amount significantly less than our simple calculation predicts. •

30–10 THE RELATIVISTIC ADDITION OF VELOCITIES

Let us recapitulate what we have learned so far in this chapter. If one observer measures the speed of sound to be $u = 340$ m/sec in still air, then another observer moving at a speed $v = 20$ m/sec will measure a sound speed $w = u + v = 360$ m/sec if he moves toward the source of the sound, and $w = u - v = 320$ m/sec if he moves away from the source. However, the Michelson-Morley experiment strongly suggests that a similar experiment with light in vacuum will yield the same speed c for both observers. The "v" neither adds to nor subtracts from the "c."

The Fizeau experiment provides us with an important link between the simple algebraic addition of low velocities and the constancy of the speed of light in vacuum. The observed time delay between the two beams is about half the delay expected on the assumption that light travels with a speed $c/n + v$ one way, and $c/n - v$ the other way. So it appears as if only about $v/2$ was added to (or subtracted from) c/n. In other words, the Fizeau experiment shows that when the light moved with the water flow its speed with respect to the observer was only about $c/n + v/2$, and when it moved against the flow its speed was about $c/n - v/2$.

Now we begin to recognize a trend: the simple addition of velocities becomes distorted as one of them approaches the speed of light. At the speed of light the addition breaks down completely.

Let us try to guess the general law of addition of velocities which holds for high as well as low velocities. We shall do this by augmenting the simple relation $w = u + v$ by a correction factor which we denote by g. Thus we want to try to write the general law of addition in the form $w = g(u + v)$. What do we know about the value of g? First of all, its value must depend on u and v, or in other words, g is a function of u and v. When both u and v are very small compared to the speed of light, g is extremely close to 1. This is just the domain in which the simple addition of velocities holds. When $u = c$, we found that $w = c$. Inserting c for u and w in the general law of addition, we find

$$c = g\,(c + v) \text{ or } g = \frac{c}{c + v} = \frac{1}{1 + v/c}$$

when $u = c$.

Now, if a boy runs with velocity v on a train moving with velocity u, or if he runs with velocity u on a train moving with velocity v, in both cases his velocity with respect to the ground is $w = u + v$. We expect this symmetry to hold also for high speeds where the simple addition fails. This means that the correction factor g must depend in the same way on both u and v. This requirement, plus the special forms g has to have when either $v = c$ or $u = c$, and when both u and v are very small, strongly suggest the following expression for g:

$$g = \frac{1}{1 + uv/c^2}$$

This expression depends in the same way on u and v and takes the desired form in the special cases mentioned before. Thus, if our guess is correct, the general way velocities add in the same direction is

$$w = \frac{u + v}{1 + uv/c^2}$$

This relation is often referred to as the relativistic addition of velocities.

Historically the relativistic addition of velocities was derived differently. Nevertheless, the method which we used here illustrates a widely used approach: When we do not know a general law, we examine special cases and look for one expression that will fit them all, and thus is likely to have general validity. This expression must then be tested experimentally. We shall use a similar approach in Section 30 - 13.

30–11 AN EXPERIMENTAL CHECK

The Fizeau experiment (Section 30 - 9) offers a natural test for the relativistic addition of velocities which we have just developed. The analysis of the experiment remains almost unchanged except that we replace $u+v$ by $(u+v)/(1+uv/c^2)$ for the speed of light traveling with the water, and we replace $u-v$ by $(u-v)/(1-uv/c^2)$ for the speed of light traveling against the water (Fig. 30 - 12). The difference in travel time of the two beams is then given by

$$\Delta t = \frac{2l\,(1-uv/c^2)}{u-v} - \frac{2l\,(1+uv/c^2)}{u+v}$$

$$= \frac{2l}{c^2}\left(\frac{(c^2-uv)(u+v)-(c^2+uv)(u-v)}{u^2-v^2}\right)$$

$$= \frac{4lv}{c^2}\left(\frac{c^2-u^2}{u^2-v^2}\right)$$

As before, v is very small compared to $u = c/n$, so we again neglect v^2 in the denominator and get

$$\Delta t = \frac{4lv}{c^2}\cdot\frac{c^2-\dfrac{c^2}{n^2}}{\dfrac{c^2}{n^2}} = \frac{4l}{c}\cdot\frac{v}{c}\,(n^2-1)$$

Let us compare this result with that found in Section 30 - 9:

$$\Delta t = \frac{4l}{c}\,\frac{v}{c}\,n^2$$

According to our new calculation, the time difference is shortened by a factor $(n^2-1)/n^2$ which for water is $(1.33^2-1)/1.33^2 = 0.44$. Thus the relativistic addition of velocities predicts a shift of $0.44 \times 0.47 = 0.21$ fringes instead of the 0.47 predicted by our previous calculation. This compares well with the observed shift of 0.23 fringes, and gives us confidence that we have found the relativistic addition of velocities.

Now, however, we must examine a deeper problem. Velocity, by definition, is displacement divided by time. If large velocities do not behave the way we know low velocities to behave, then we become suspicious that our notions about length and time may not be adequate. We shall have to examine carefully what we really mean when we measure intervals of length and time in a frame of reference which is moving with respect to us.

30–12 TIME FOR DIFFERENT OBSERVERS

Let us consider the motion of a body at point P as viewed in two frames of reference moving with respect to each other with velocity $\vec{v}$. To be specific, we refer to the frame in which every point has the coordinates x, y, z as the fixed frame, and the other frame, with coordinates x', y', z', as the moving frame. Furthermore, it is convenient to have the axes of the two frames parallel to each other, and to have the direction of the relative motion along the x axis (Fig. 30 - 13). It is evident from Fig. 30 - 13 that the coordinates of point P on the x axis in the two frames of reference are related by the equation

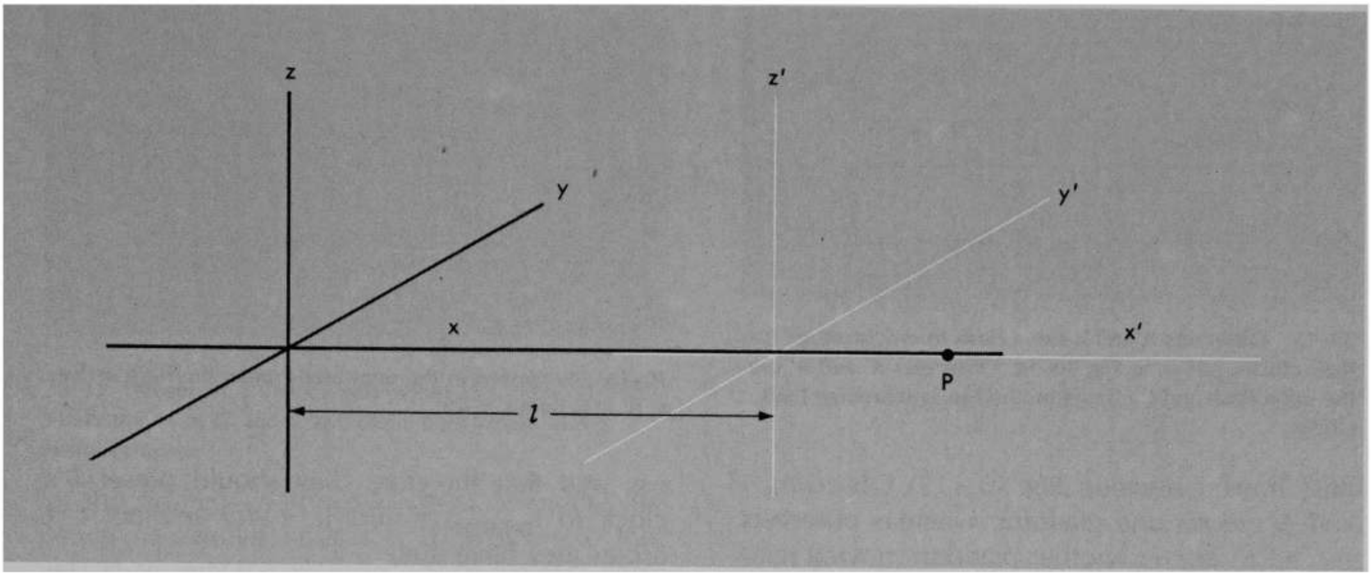
$$x = x' + l$$

For further convenience, we choose the moment when the two coordinate systems coincide as the starting point of our time scale; that is, $t = 0$. What we mean is that observers at the origin in both frames of reference, equipped with identical stop watches, will start them off the instant the axes of their frames of reference coincide. Then the distance between the origins of the two frames is expressed by $l = vt$, and at any time t, the coordinates of P in the two frames are related by $x = x' + vt$. In particular, at times t_1 and t_2 we have

$$x_1 = x'_1 + vt_1$$
$$x_2 = x'_2 + vt_2$$

Subtracting the first equation from the second and dividing by $t_2 - t_1$ yields

$$\frac{x_2-x_1}{t_2-t_1} = v + \frac{x'_2-x'_1}{t_2-t_1}$$



30–13 The motion of a body at P can be viewed either in the unprimed frame of reference (black axes) or the primed frame (white axes), which moves to the right. At the instant shown the origins of the two frames are a distance l apart.

We can write this as $\Delta x/\Delta t = v + \Delta x'/\Delta t$, which for small enough Δt becomes

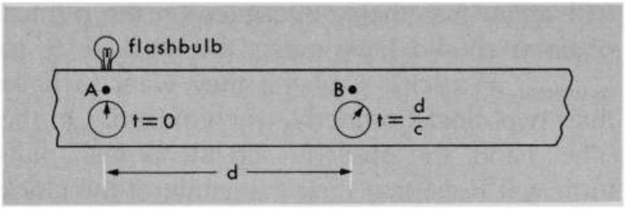
$$u = v + u'$$

Now, this is precisely the simple addition of velocities which was found to fail at high speeds. We must have overlooked something fundamental in this simple argument.

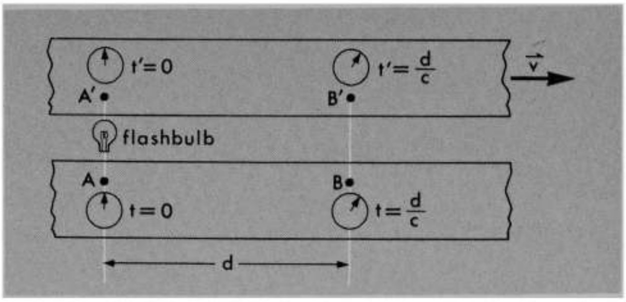
Looking at Fig. 30 – 13, it is hard to find fault with the relation $x = x' + l$, but are we sure that the time is the same in both frames of reference? As so often happens in science, asking the right question is what requires the great insight. Once the right question has been asked, the answer may be simple. The question we have just raised was asked by Albert Einstein in 1905. Let us study it in detail.

Measuring time requires a clock. Reading times simultaneously at two locations requires two synchronized clocks, that is, two clocks which agreed at some given instant and which run at the same rate.

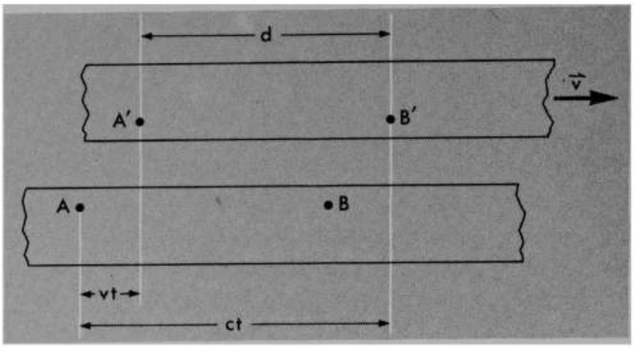
We have seen in Section 30 – 3 that we can use a light signal to synchronize two clocks that are at different positions only if we assume that the speed of light is the same in all directions. By now we have presented a good deal of evidence that in vacuum this is indeed the case. Thus we can synchronize two clocks a distance d apart by firing off a flash halfway between them and have the two observers start their clocks when they see the flash. An even simpler procedure is the following: We place the flashbulb at A as in Fig. 30 – 14 instead of halfway between A and B. Observer A presets his clock to $t = 0$ while observer B presets his to $t = d/c$ to allow for the time required by the flash to travel from A to B. The flash starts A's clock and upon arrival at B starts the latter's clock. Since the speed of light is a universal constant, the same procedure can be used by two other observers in any other frame of reference, moving with respect to the first one with a constant velocity v. Now, there is no reason why two pairs of observers in different frames of reference could not use the same flash to synchronize their clocks. You can imagine



30–14 Synchronizing two clocks by use of a flashbulb. The clocks are preset as shown. The clock at A starts when the flash goes off; the clock at B starts when the flash reaches it.



30–15 Observers A and B use a flash to synchronize their clocks, just as in Fig. 30–14. Observers A' and B' use the same flash and the same method to synchronize their clocks.



30–16 As viewed in the unprimed frame, the flash arrives at B' at a time t > d/c.

the following situation (Fig. 30 – 15). Observers *A* and *B* are on one platform, whereas observers *A'* and *B'* are on another platform moving relative to the first with velocity *v*. When *A'* passes *A* a flash is set off and both pairs of observers use it to synchronize their clocks. Observers *A* and *B* will be satisfied that their clocks are synchronized, and so will *A'* and *B'*, but what will each pair say about the other pair's clocks?

Let us calculate how the synchronization experiment in the primed frame of reference will appear to the observers in the unprimed frame. Since the platform in the primed system moves to the right relative to the unprimed system, it will take more time for the light to reach *B'* than to reach *B* (Fig. 30 – 16). Call *t* the time needed for the light to get from *A* to *B'*; then during this time the primed platform has moved a distance *ut* and the light flash a distance *ct*. Their difference is $ct - vt = d$, and

$$t = d \cdot \frac{1}{c - v} = \frac{d}{c} \cdot \frac{1}{1 - v/c}$$

To the first order in v/c this yields

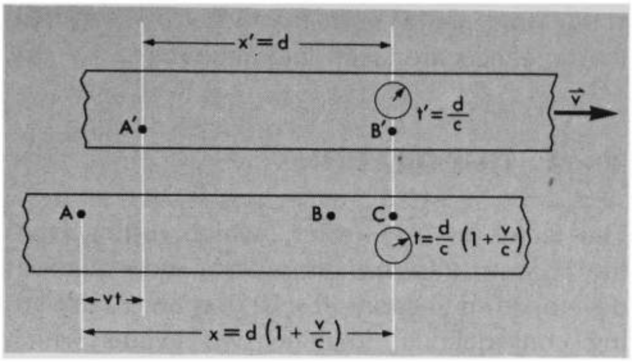
$$t = \frac{d}{c}\left(1 + \frac{v}{c}\right)$$

Thus the observers on the unprimed platform will argue that their colleagues on the primed platform should have preset the clock of *B'* to $t'_{\text{apparent}} = (d/c)(1 + v/c)$ if they want to have their two clocks properly synchronized. On the other hand, the observers on the primed platform will insist that their presetting of the clock to $t' = d/c$ is correct, because as far as they are concerned they are not moving at all. Moreover, the observers in the primed platform will justly claim that the unprimed platform and its observers are moving to the left with velocity $-v$, and that therefore they should preset *B*'s clock to $t_{\text{apparent}} = (d/c)(1 - v/c)$ and not $t = d/c$ as they have done.

This disagreement has far-reaching consequences. We use synchronized clocks to register the time when a certain event takes place. For example, with two synchronized clocks we can tell when two stones were dropped at two points along a railroad track. If the two time readings are identical we say that the two events took place simultaneously. From our preceding discussion it follows that if observers at rest on the ground claim that the two stones were dropped simultaneously, then persons on a train passing by will challenge this conclusion because, according to their clocks, the clocks on the ground are not synchronized. Thus, in two frames of reference moving with respect to each other, the time scales differ. How can we relate them?

Let us go back to Fig. 30 – 15. Observers *A* and *A'* preset their clocks to zero, while the clocks of *B* and *B'* are preset to d/c. According to *B*, observer *B'* should have preset his clock (to the first order in v/c) to $(d/c)(1 + v/c)$. This is the time to which a third observer *C* located at a distance $x = d\ (1 + v/c)$ from *A* in the unprimed frame would have preset his clock in order to have it synchronized by the same light flash that was used by *B* and *B'* (Fig. 30 – 17). When *B'* and *C* receive the light flash and start their clocks, that of *B'* will start from $t' = d/c$, and that of *C* from $t = d/c + dv/c^2$. If both clocks are made in the same way, we expect them to maintain the same difference from then on. Thus the times on these two clocks will always be related by $t = t' + dv/c^2$.

Observer *B'* can be placed at any distance from *A'*, whose location can serve as the origin


$x' = d$
A'
B'
$t' = \frac{d}{c}$
$\vec{v}$
A
B
C
$t = \frac{d}{c}\left(1 + \frac{v}{c}\right)$
vt
$x = d\left(1 + \frac{v}{c}\right)$

30–17 Observer C's clock was preset to t = (d/c) (1 + v/c). Flash arrives at B′ and C and starts both clocks at the instant shown.

of the primed coordinate system. Thus the distance d is quite arbitrary. To emphasize this, we shall replace d by x'. Furthermore, since the clocks of A, B, and C are synchronized, the relation $t = t' + vx'/c^2$ establishes the relation between all clocks synchronized in the unprimed frame of reference and a clock located at a distance x' from the origin in the primed frame of reference.

From Fig. 30 – 17 it follows that $x = x' + vt$. To express x in terms of the coordinate and time in the primed frame of reference we can substitute t from the preceding equation:

$$x = x' + vt' + x' \frac{v^2}{c^2}$$

Since so far we carry our calculations only to first order in v/c, we shall neglect the last term and write

$$x = x' + vt'$$

This relation is obvious from the point of view of an observer in the primed frame of reference: Relative to him the unprimed frame has moved a distance vt' to the left; hence, the unprimed coordinate is that much larger than his primed coordinate. Thus we learned to describe events in terms of two sets of coordinates, (x, t) and (x', t'), belonging to two frames of reference moving with relative velocity v. They are related by the transformation equations

$$x = x' + vt'$$

$$t = t' + \frac{vx'}{c^2}$$

We can now check whether these equations can lead to the relativistic addition of velocities. The velocity of an object moving along the x axis and measured in the unprimed system is $u = \Delta x/\Delta t$. In the primed system, it is $u' = \Delta x'/\Delta t'$. But

$$\Delta x = \Delta x' + v\Delta t'$$

and

$$\Delta t = \Delta t' + v\Delta x'/c^2$$

By dividing the first equation by the second, we find

$$u = \frac{\Delta x}{\Delta t} = \frac{\Delta x' + v\Delta t'}{\Delta t' + \dfrac{v\Delta x'}{c^2}}$$

$$= \frac{\dfrac{\Delta x'}{\Delta t'} + v}{1 + \dfrac{v}{c^2} \cdot \dfrac{\Delta x'}{\Delta t'}} = \frac{u' + v}{1 + \dfrac{vu'}{c^2}}$$

This is the relativistic addition of velocities which has already been checked experimentally.

Note that before we looked into the matter of simultaneity, we intuitively assumed $t = t'$, and therefore did not bother to introduce t' at all. This, in turn, leads only to the nonrelativistic addition of velocities.

30–13 THE LORENTZ TRANSFORMATION

Our derivation of the equation $t = t' + ux'/c^2$ was restricted to the case that v is much smaller than c, since we carried our calculation only to first order in v/c. To find out whether our results hold also when this condition is not satisfied, we shall check the consistency of our set of equations expressing x and t in terms of x' and t', and those expressing x' and t' in terms of x and t. Since the primed frame of reference is moving with velocity v with respect to the unprimed frame, the latter moves with a velocity $-v$ with respect to the former. Thus, by replacing v by $-v$ and x by x', and t by t', we get the set of equations expressing x' and t' in terms of x and t:

$$x' = x - vt$$

$$t' = t - ux/c^2$$

If these two sets of equations are consistent with each other, then by substituting the expressions for x' and t' from the second set of equations into the first set, we should end up with the identities $x = x$ and $t = t$. Instead we find

$$x = x' + vt' = x - vt + vt - v^2/c^2\, x$$
$$= (1 - v^2/c^2)x$$
$$t = t' + vx'/c^2 = t - vx/c^2 + vx/c^2 - v^2t/c^2$$
$$= (1 - v^2/c^2)t$$

The factor $(1 - v^2/c^2)$ contains a second-order term in v/c but not a first-order term. This is encouraging. It shows that the transformations

$$x = x' + vt$$
$$t = t' + vx'/c^2$$

can be applied as long as $1 - v^2/c^2 \approx 1$. This includes situations where the nonrelativistic relations $x = x' + vt'$ and $t = t'$ failed completely. We must now find a way of extending our transformation to larger values of v/c without destroying their validity in the region of smaller values of v/c. We can proceed as follows. Since we ended up with the factor $(1 - v^2/c^2)$ after applying our approximate transformation twice, we can eliminate this factor entirely by multiplying the right-hand side of each set of equations by the factor $1/\sqrt{1 - v^2/c^2}$; then, by performing the substitution in the consistency check, we shall have a factor $1/(1 - v^2/c^2)$, which will cancel the undesirable $(1 - v^2/c^2)$. The relations then become

$$x = \frac{1}{\sqrt{1 - v^2/c^2}}\,(x' + vt')$$

$$x' = \frac{1}{\sqrt{1 - v^2/c^2}}\,(x - vt)$$

$$t = \frac{1}{\sqrt{1 - v^2/c^2}}\,(t' + vx'/c^2)$$

$$t' = \frac{1}{\sqrt{1 - v^2/c^2}}\,(t - vx/c^2)$$

This transformation of coordinates is known as the Lorentz transformation. It provides the central pillar of the special theory of relativity, but before we trust it we must find an experimental check for the validity of the new factor $1/\sqrt{1 - v^2/c^2}$.

What are the relations between y, z and y', z'? Figure 30 – 13 suggests that both $y = y'$ and $z = z'$, independent of x', t' and v. Furthermore, all our considerations which led to the Lorentz transformation would remain unchanged if the x and x' axes were displaced parallel to themselves. Therefore, the relations $y = y'$ and $z = z'$ remain valid also when relativistic effects are taken into account.

30–14 TIME DILATION

The factor $1\ \sqrt{1 - v^2/c^2}$, which distinguishes the Lorentz transformation from the equations developed in Section 30 – 12, has a very intriguing consequence. Imagine two events which occur at the same location, say at x', in the moving (primed) frame of reference, one at a time t'_1 and the other at a time $t'_2 = t'_1 + \Delta t'$. The corresponding times for an observer in the fixed (unprimed) frame of reference are t_1 and $t_2 = t_1 + \Delta t$. How is the time difference $\Delta t'$ related to Δt? From the Lorentz transformation, we find

$$t_1 = \frac{1}{\sqrt{1 - v^2/c^2}}\,(t'_1 + vx'/c^2)$$

$$t_2 = \frac{1}{\sqrt{1 - v^2/c^2}}\,(t'_1 + \Delta t' + vx'/c^2)$$

Therefore,

$$\Delta t = t_2 - t_1 = \frac{\Delta t'}{\sqrt{1 - v^2/c^2}}$$

Thus, the observer in the fixed system, on comparing notes with the observer in the moving (primed) system, will claim that the time interval between the two events is larger by the factor $1/\sqrt{1 - v^2/c^2}$. The nature of the events is quite irrelevant. They can be the emission of two pulses, the radioactive disintegration of two atoms, or successive arrivals of a pendulum at one end of its swing. The time interval between any of these pairs of events at the same place in the moving frame of reference will appear longer to an observer in the fixed frame, even though both observers are using identical clocks. The moving clock appears to him to go slower; this effect is called *time dilation*.

At this point, we must remember that our designation of one system as fixed and one as moving was made for the convenience of expression. Both systems are moving relative to each other with speed v. We could equally well refer to the primed system as the "fixed" system and to the unprimed one as the "moving" system. Then, following the same argument as before, we find that the time interval Δt between two events which occur at the same place in the unprimed system appears to the observer in the primed system as $\Delta t'$, where now $\Delta t' = \Delta t/\sqrt{1 - v^2/c^2}$.

Can both relations

$$\Delta t = \frac{\Delta t'}{\sqrt{1 - v^2/c^2}} \quad \text{and} \quad \Delta t' = \frac{\Delta t}{\sqrt{1 - v^2/c^2}}$$

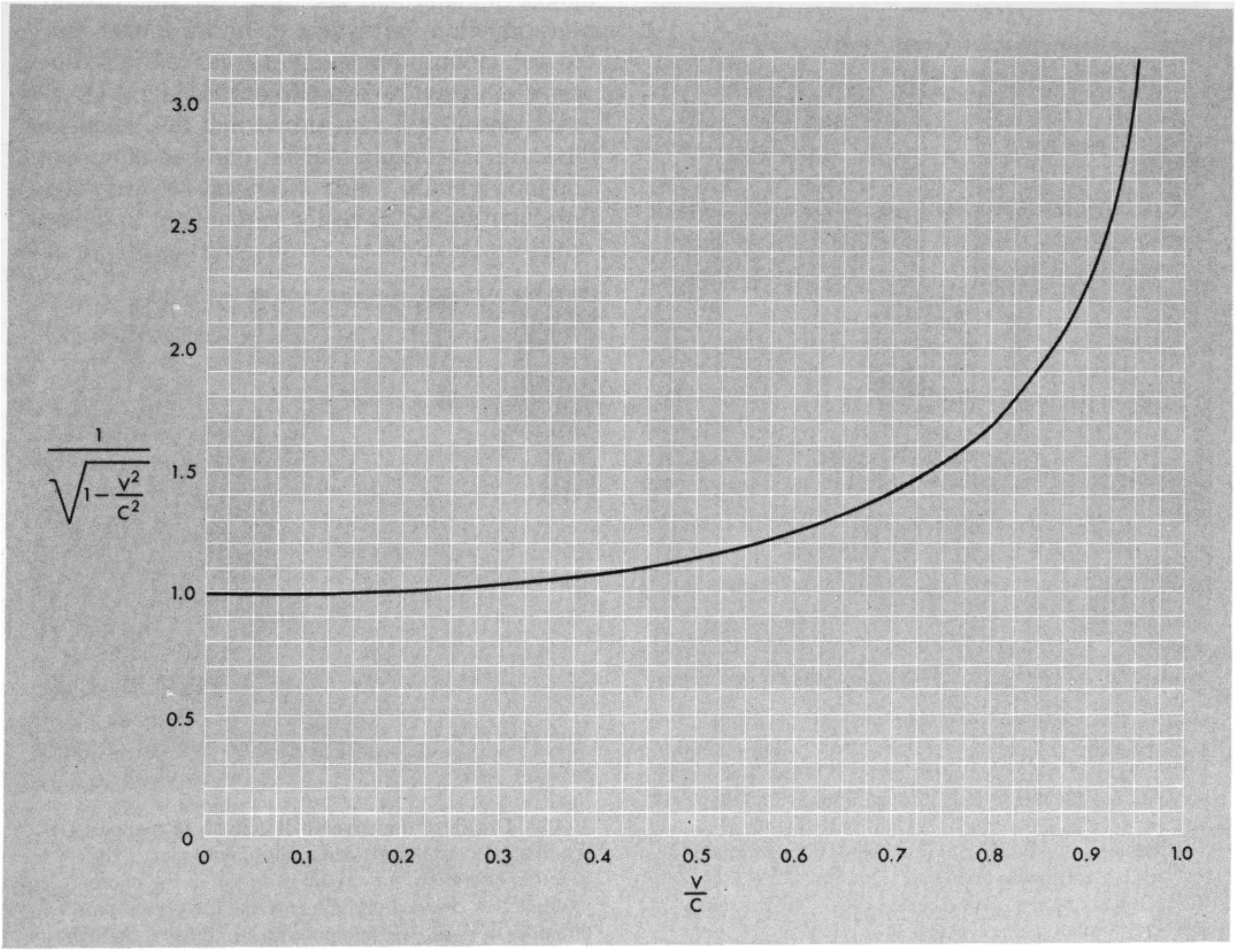
be correct, or should the second one read

$$\Delta t' = \Delta t \sqrt{1 - v^2/c^2}?$$

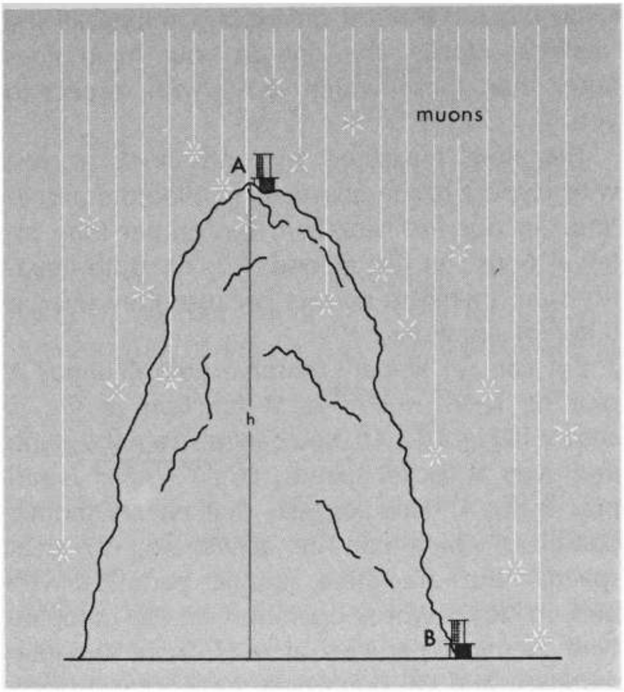
To answer this question, let us see *where* the two events in question occur. In the first case, they occurred at the same point, x', in the primed system, and at different points, separated by a distance $x_2 - x_1 = v\Delta t$ in the unprimed system. On the other hand, in the second case, the two events occurred at the same point in the unprimed system and, therefore, at different positions as seen by an observer in the primed system. The two cases represent different situations. In each the time interval is shortest for the observer for whom the events occur at the same place. To put it differently, among all the identical clocks, the one in your hand goes faster than those which move with respect to you.

The time measured by one clock at rest with respect to the observer is called the *proper time*. In our first situation, the proper time interval is $\Delta t'$; in the second, it is Δt. Both equations are therefore correct because they refer to different situations.

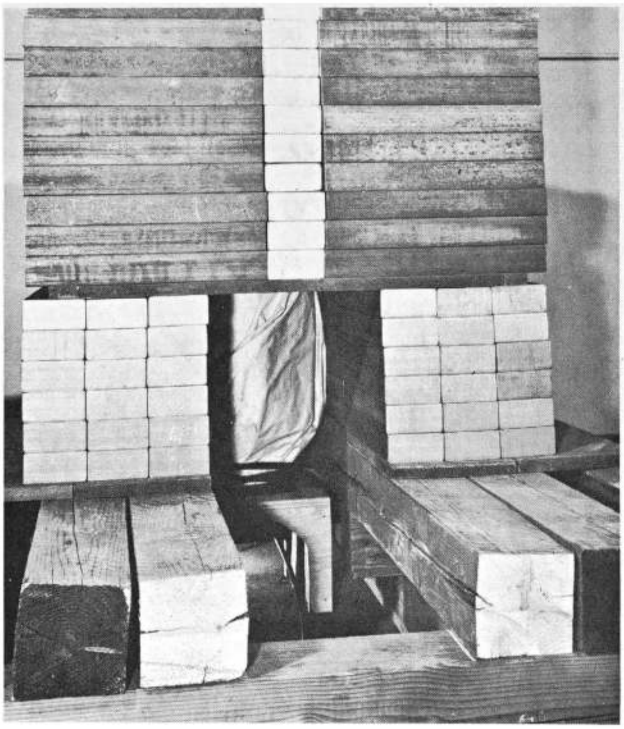
But can we actually observe time dilation? A plot of $1/\sqrt{1 - v^2/c^2}$ as a function of v/c is shown in Fig. 30-18. You can see from the graph that even at rocket speeds, $1/\sqrt{1 - v^2/c^2}$ is still practically 1. This suggests that we try atomic particles, which can be accelerated to great speeds. But are there atomic particles with built-in clocks whose operation we can compare with identical particles at rest? There *are* such particles, and we don't even have to accelerate them ourselves: they come through the atmosphere as part of the cosmic radiation.



30–18 A plot of $1/\sqrt{1 - v^2/c^2}$ vs. v/c. The speed of the fastest rocket is practically at v/c = 0 on this graph.



30–19 **Muons coming from the upper atmosphere and decaying in flight. Fewer muons reach sea level (B) than the top of the mountain (A).**

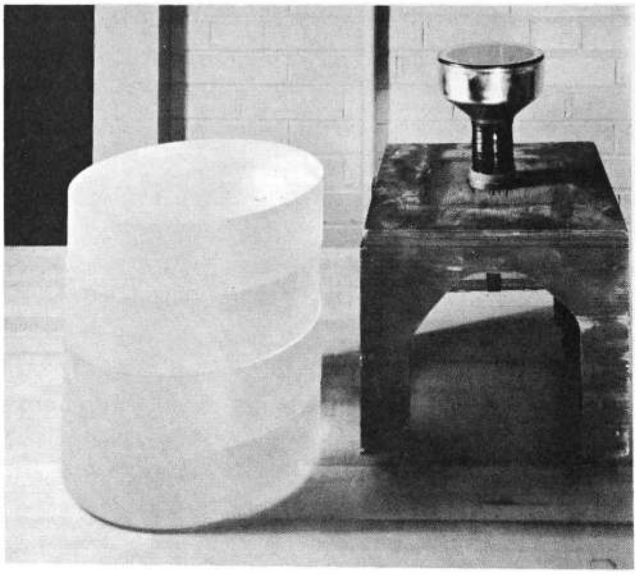


30–20 **The counter covered with opaque paper and surrounded by a shield of iron slabs. Most muons entering the shield from above are either stopped in the iron before reaching the counter or pass through both and continue into the ground. Only muons within a narrow energy band are stopped in the counter and decay. A stopped muon is, therefore, identified by two pulses in rapid succession: the first made by the muon and the second by the decay electron. A muon going through the counter makes only one pulse.**

30–15 AN EXPERIMENT WITH MUONS

Mu-mesons, or muons for short, are particles of about 200 electron masses which are produced in the upper atmosphere by the incoming cosmic rays. The muons are unstable particles which decay after a short time into an electron and two neutral particles. We cannot predict how long any individual muon will "live" before it decays, but we can measure the lifetime of many muons and find out how long it takes for any given fraction of them to decay. In particular, we can determine very accurately the time during which one half of a large number of muons would decay. This time, called the half-life, is $T = 1.5 \times 10^{-6}$ sec. Any large number of muons yields the same value for the half-life. It is the existence of the half-life of a large collection of muons which provides us with the built-in clock.

The half-life of 1.5×10^{-6} sec was obtained from measurements made on muons at rest, or almost at rest. On the basis of the Lorentz transformation we expect to find a longer half-life for muons moving at speeds close to the speed of light. An experiment done by D. H. Frisch and J. H. Smith shows that this is indeed the case. They counted the number of muons within a certain range of energy at two locations: *A*, on top of a high mountain, and *B*, a



30–21 **Details of the counter. It consists of two parts: a scintillating crystal (left) and a photomultiplier (right). When in operation, the crystal is on top of the photomultiplier. A charged particle entering the crystal produces a faint light flash. The photomultiplier converts the light signal into a sizable electric pulse.**

distance h below, at sea level (Fig. 30 – 19). The counting apparatus is shown in Figs. 30 – 20 to 30 – 22. All the muons selected by the counter have very closely the same speed. If we measured their speeds directly, we would find them to equal almost the speed of light. (See Chapter 32.) Therefore, it takes the muons a time h/c to travel from the top of the mountain (A) to sea level (B). What fraction of the muons that arrive at A will reach B? Figure 30 – 23 shows this fraction as a function of time, expressed in half-lives. Frisch and Smith chose the point A on the top of Mount Washington, and B in Cambridge, Massachusetts, so that $h = 1.9 \times 10^3$ meters and the time of flight is

$$\frac{h}{c} = \frac{1.9 \times 10^3}{3.0 \times 10^8} = 6.3 \times 10^{-6} \text{ sec}$$

If the half-life of the muon were the same at high speed and at rest, then the time of flight would amount to

$$n = \frac{h}{cT'} = \frac{6.3 \times 10^{-6}}{1.5 \times 10^{-6}} = 4.2 \text{ half-lives}$$

From Fig. 30 – 23 we would expect only 0.05 of the muons present on Mount Washington to reach sea level. On the other hand, if a time dilation really exists, the half-life of the muons will appear to us to be $T = T'/\sqrt{1 - v^2/c^2}$ and, therefore, the time required from the height of Mount Washington down to sea level amounts to only

$$\begin{aligned} n &= h/cT' \sqrt{1 - v^2/c^2} \\ &= 4.2 \sqrt{1 - v^2/c^2} \text{ half-lives} \end{aligned}$$

The actual fraction observed with the apparatus shown in Fig. 30 – 20 to 30 – 22 was 0.68. This corresponds to $n = 0.46$ half-life in Fig. 30 – 23. Hence,

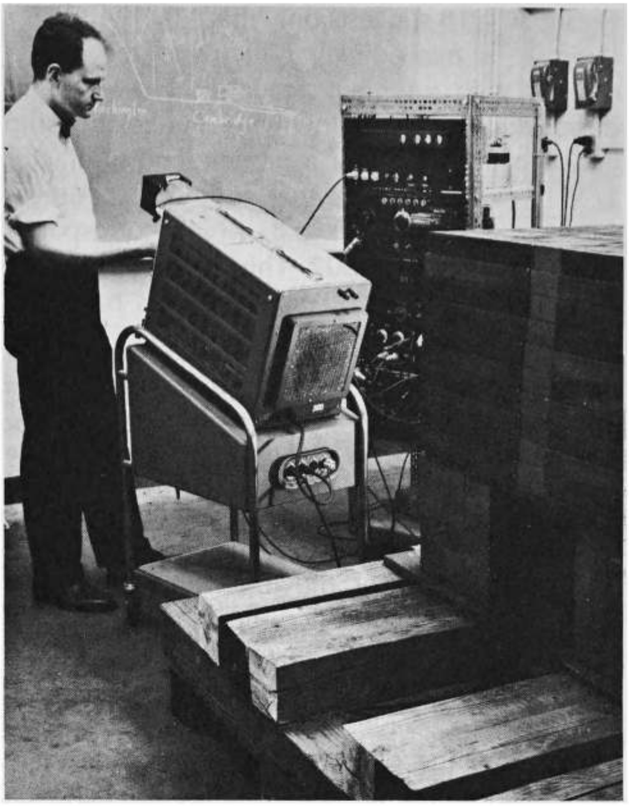
$$0.46 = 4.2 \sqrt{1 - v^2/c^2}$$

or

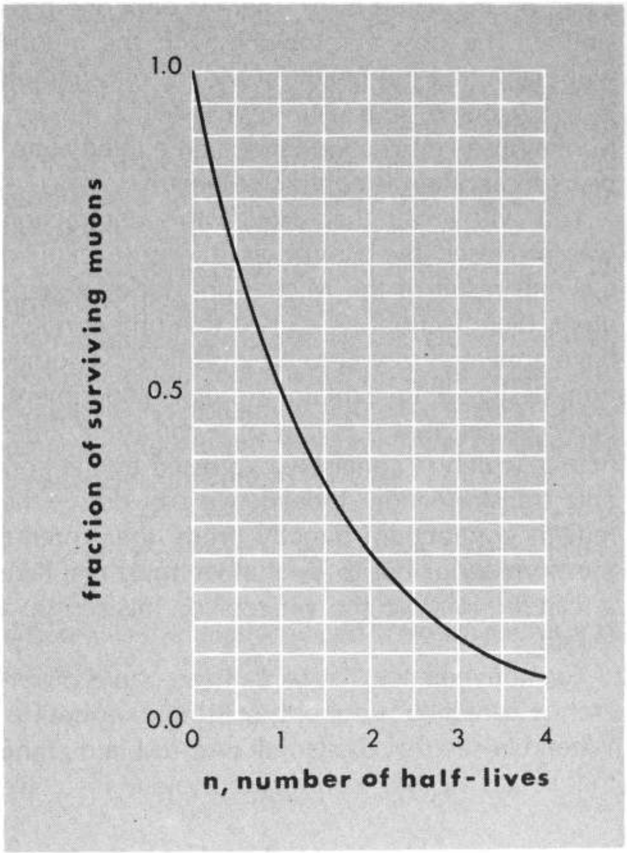
$$\frac{1}{\sqrt{1 - v^2/c^2}} \approx 9$$

We interpret this result as indicating that the muon clocks moving at high speed relative to us run at 1/9 the rate of muon clocks at rest in our frame of reference.

This experiment certainly demonstrates the time dilation, although it cannot test the relationship $T = T'/\sqrt{1 - v^2/c^2}$ quantitatively without some independent determination of v.



30–22 General view of the apparatus. The registering equipment on the panel counts only muons that are stopped in the scintillating crystal. The pulses are selected by the oscilloscope and relayed to the scaler by a second photomultiplier.



30–23 The characteristic decay curve ($v = e^{-x}$). Here it represents the fraction of muons which have not decayed as a function of time expressed in terms of the number of muon half-lives.

Figure 30 – 18 suggests only that $1/\sqrt{1 - v^2/c^2}$ is large. It is exceedingly difficult to measure v directly with sufficient accuracy. However, indirect measurements involving the relationship between the energy of a particle and the factor $1/\sqrt{1 - v^2/c^2}$ indicate that in this experiment $1/\sqrt{1 - v^2/c^2} = 8.8 \pm 0.8$, which corresponds to $v/c = 0.994 \pm 0.002$. This confirms the prediction of the time dilation which is based on the Lorentz transformation.

30–16 THE LENGTH OF MOVING OBJECTS

To us, the half-life of muons rushing down the atmosphere appears extended. How will it appear to an observer moving with the muons? How will he account for the large fraction of muons which succeed in making the trip from the height of Mount Washington to sea level?

Since the observer at rest with respect to the muons and the observer on the ground count the same number of particles, they will agree on the value of $n = (h/cT')\ \sqrt{1 - v^2/c^2}$. They will also agree on their relative velocity v. The observer on the earth will say that the half-life of the muons is $T'/\sqrt{1 - v^2/c^2}$; the other observer will claim that it is T'. How, then, can they agree on the value of n? There is only one possibility. The observer moving with the muons will claim that the height of Mount Washington is not h but $h' = h\sqrt{1 - v^2/c^2}$. What appears to one observer as an extended time interval appears to another as contracted length.

You will recall that exactly this contraction was proposed by Lorentz as a way to explain the null result of the Michelson-Morley experiment. (See Section 30 – 5.) At that time it might have appeared to you to be an isolated assumption designed to explain a single experiment. Now we see that it is part of a consistent kinematics which is concisely expressed by the Lorentz transformation. Indeed, we can derive the length contraction directly from the Lorentz transformation; but to do that we must first take a careful look at the process of measuring a length.

For convenience, let us call the frame of reference in which we are at rest the fixed frame. When we say that the length of a rod laid along the x axis of our coordinate system is L, we mean that at a certain moment, at t, say, the two ends of the rod had the coordinates x_1 and $x_2 = x_1 + L$. How would an observer in the moving frame measure the length of the same rod? Using the same approach as we do, he will say that the length, L', of the rod is the difference, $x'_2 - x'_1$, between the coordinates of the ends of the rod on his x' axis at a certain instant as shown by his clocks, that is, at a time t'. Now, we know that the same time t' at the two locations x'_1 and $x'_2 = x'_1 + L'$ corresponds to different times t. Thus; $L' = x'_2 - x'_1$

$$= \frac{1}{\sqrt{1 - v^2/c^2}}\,(x_2 - vt_2)$$

$$- \frac{1}{\sqrt{1 - v^2/c^2}}\,(x_1 - vt_1)$$

$$= \frac{1}{\sqrt{1 - v^2/c^2}}\,[x_2 - x_1 - v(t_2 - t_1)]$$

The time difference $(t_2 - t_1)$ is obtained from the transformation equation for t':

$$t' = \frac{1}{\sqrt{1 - v^2/c^2}}\,(t_2 - vx_2/c^2)$$

$$= \frac{1}{\sqrt{1 - v^2/c^2}}\,(t_1 - ux_1/c^2)$$

Hence,

$$t_2 - t_1 = (v/c^2)\,(x_2 - x_1)$$

Substituting this expression into the equation for L', and recalling that $x_2 - x_1 = L$, we find

$$L' = \frac{L}{\sqrt{1 - v^2/c^2}}\,(1 - v^2/c^2) = L\sqrt{1 - v^2/c^2}$$

which is the length contraction we were looking for.

30–17 SUMMARY

We started this chapter with a search for the light-carrying medium—the "ether." Our objective seemed rather modest: all we wanted was to identify the frame of reference in which the ether is at rest. The experiments which were to give us the answer were based on a wealth of experience with waves of all kinds, yet they failed to show the expected results. As happens often in physics, the failure provoked a fresh examination of all the elements which went into the theory, and, in particular, a careful look at all the extrapolations which were made.

Contrary to what you may have expected, the trouble was not with the wave model of light. Our basic ideas on time and space needed refinement. We found that the universality of the speed of light plays a central role in kinematics, because we need light to transmit signals which tell us when things happen and how our clocks run. This is how the speed of light enters into the transformation of the time coordinates, and into the relativistic addition of velocities.

Finally, to fit all the pieces into a consistent scheme, we had to introduce the factor $1/\sqrt{1-v^2/c^2}$ into the transformation equations. This factor does not make any sense when $v > c$. We can, of course, imagine an abstract frame of reference to move at a speed $v > c$ with respect to us. But can we actually make a material body move with a speed $v > c$? According to Newtonian dynamics, this should be entirely possible. What really happens is the subject of the next chapter.

APPENDIX: THE TWIN PARADOX

You may have heard about the "twin paradox": one of two twins rockets away from the earth for a certain distance, then turns around and rockets back. To him it appears that the earth and the other twin made a round trip in the opposite direction. Will they both age the same amount? Will each claim that the other aged less because of time dilation? Or will one age less than the other?

Assume that twin A stays on the earth while twin B leaves the earth at speed v and travels a distance d as measured in the earth frame; then he comes to a sudden stop and returns to earth, again with speed v. During the whole trip they broadcast their heartbeats to each other. Because they are identical twins, their hearts beat at the same frequency f when on the earth.

In what follows we shall neglect any effects of the starting, turning, and stopping accelerations on the heartbeat of B. Furthermore, we shall assume that almost no time was spent accelerating, so that B was essentially moving half the time away from the earth and half the time toward the earth, all with speed v.

In A's frame, the outward and inward journeys each last a time $t = d/v$. For B the distance is Lorentz-contracted and appears to be $d\sqrt{1-v^2/c^2}$. Hence each journey takes him only a time

$$t' = (d/v)\ \sqrt{1-v^2/c^2}$$

The total journey lasts a time $2d/v$ for A, during which he counts $f \cdot 2d/v$ heartbeats for himself. For B, the total journey lasts a time $(2d/v)\sqrt{1-v^2/c^2}$, during which he counts $f \cdot (2d/v\sqrt{1-v^2/c^2}$ heartbeats for himself. Thus twin A counts more heartbeats for A during the trip than B does for B.

How many heartbeats does each count for the other from the broadcast of the heartbeats? We must recognize the fact that the frequency received differs from the frequency emitted, owing to the relative motion of source and observer, a phenomenon called the *Doppler effect.*

Even for sound and water waves, this phenomenon occurs. Indeed, Doppler effects in sound are quite familiar in everyday experience. If source and observer are separating, the apparent frequency is lower than the actual frequency; if they are approaching each other, the apparent frequency is higher. (See Questions 2 and 3.) For light (or radio) signals, the effects are similar. Since lower frequency corresponds to longer wavelength and higher frequency to shorter wavelength, it has become the usual terminology to refer to the two cases as the "red shift" and "blue shift," respectively.

Because the speed of light is the same for all observers moving relative to one another, the quantitative expressions for the Doppler-shifted frequencies are not the same as for sound and water waves. Instead, as you have seen in Questions 39 and 40, the red-shifted and blue-shifted frequencies turn out to be

$$f_{\text{red}} = f\sqrt{\frac{1-v/c}{1+v/c}} \quad \text{and} \quad f_{\text{blue}} = f\sqrt{\frac{1+v/c}{1-v/c}}$$

where v is the relative speed between source and observer.

During the outward journey, B receives a red-shifted signal from A for a time

$$t'_{\text{red}} = (d/v)\sqrt{1-v^2/c^2}$$

The instant he reverses direction, the red shift turns to a blue shift. During the inward journey, B receives a blue-shifted signal for a time

$$t'_{\text{blue}} = (d/v)\sqrt{1 - v^2/c^2}$$

Twin A also receives a red-shifted signal from B during the outward journey. When B reverses direction, A continues to receive the red-shifted signal for an additional time d/c, the time it takes for signals to reach him when B is farthest away. Thus A will receive a red-shifted signal for a total time

$$t_{\text{red}} = (d/v + d/c) = (d/v)(1 + v/c)$$

and a blue-shifted signal for the remaining time,

$$t_{\text{blue}} = (d/v - d/c) = (d/v)(1 - v/c)$$

It is this time delay which is the essential asymmetry of the problem.

The total number of B's heartbeats counted by A is

$$\begin{aligned} &f_{\text{red}}t_{\text{red}} + f_{\text{blue}}t_{\text{blue}} \\ &= f\sqrt{\frac{1 - v/c}{1 + v/c}} \cdot \frac{d}{v}(1 + v/c) \\ &\qquad + f\sqrt{\frac{1 - v/c}{1 + v/c}} \cdot \frac{d}{v}(1 - v/c) \\ &= f \cdot \frac{2d}{v}\sqrt{1 - v^2/c^2} \end{aligned}$$

which is exactly what B counted for himself.

The total number of A's heartbeats counted by B is

$$\begin{aligned} &f_{\text{red}}t'_{\text{red}} + f_{\text{blue}}t'_{\text{blue}} \\ &= f\sqrt{\frac{1 - v/c}{1 + v/c}} \cdot \frac{d}{v}\sqrt{1 - v^2/c^2} \\ &\qquad + f\sqrt{\frac{1 + v/c}{1 - v/c}} \cdot \frac{d}{v}\sqrt{1 - v^2/c^2} \\ &= f \cdot \frac{2d}{v} \end{aligned}$$

which is exactly what A counted for himself.

Thus, both twins are in complete agreement as to the total number of heartbeats each has had during B's trip, and both conclude that B, having had fewer heartbeats, is younger. There is really no paradox.

QUESTIONS

1.* In the situation shown in Fig. 30 – 1, what values do t_{AB} and t_{BA} approach when the speed of the water approaches the speed of the waves?

2. A straight-wave generator in a swimming pool emits waves of period T and velocity u.
   (a) What is the wavelength of the waves?
   (b) These waves are detected by a swimmer moving at a speed v toward the source. What is the time interval T' between successive crests passing the swimmer?

3. The source in Question 2 now moves through the water at speed v toward the swimmer, who is treading water.
   (a) What is the wavelength λ' of the waves observed by the swimmer?
   (b) What is the period T' observed by the stationary swimmer?

4. Both source and swimmer in Questions 2 and 3 move in the same direction, at the same speed v with respect to the water, while a second swimmer near the first does not move relative to the water. Assume that the source moves toward both swimmers.
   (a) What is the period T' for the swimmer at rest in the water?
   (b) What is the time T'' between successive crests observed by the moving swimmer?

5. Galileo made an attempt to measure the speed of light by uncovering a lantern on the top of one mountain while an assistant uncovered a second lantern on another mountain 1 mile away when he saw Galileo's light. Considering the reaction time of the assistant, what could Galileo conclude from his experiment?

6.* In Fig. 30 – 2, suppose $v/u = 0.5$. At what angle to the bank would you send a pulse from A in order that it will arrive at C?

7.* Expand $\dfrac{\sqrt{1 + x^2}}{1 - x}$ to the second order in x.

8. Use the power expansion to the second order to find an *approximate* numerical value for the quantity $\dfrac{1}{\sqrt{1 - v^2/c^2}}$ corresponding to the following values of v:
   (a) $0.2c$, (b) $0.4c$, (c) $0.6c$, (d) $0.8c$, (e) $1.0c$.
   Compare your values with the exact ones shown in Fig. 30 – 18.

9. (a) Expand the expression $\frac{1}{1-x}$ to include the fifth-order term.

(b) If $x = 0.0200$, how many terms must be included in the expression you have found?

(c) If $x = 0.200$, how many terms must be included?

10. Compare the expansion of T' in terms of T in Questions 2 and 3 to first and second order in v/u.

11. If the ether is at rest with respect to the sun, with approximately what speed would the earth have to move around the sun in order for there to be an observable shift in the interference pattern seen with the interferometer described in the laboratory guide?

12. Consider a Michelson interferometer with arms of different lengths held in a fixed position. Does the interference pattern depend on the motion of the earth even if the Lorentz contraction is involved?

13.* In Fig. 30 – 5, where is the speed of the waves relative to the tank the greatest?

14. Two crests are started simultaneously at points S_1 and S_2, a distance 10 cm apart.

The wave speed is 1 cm/sec, and the medium moves from S_1 toward S_2 at a speed 0.25 cm/sec.

(a) When do the crests first touch?

(b) Draw a diagram to scale which shows the location of the double crest at 1-sec intervals during the 5 sec after the crests first touch.

(c) What is the average speed of the double crest during each of these 1-sec intervals?

15. Two stones are dropped simultaneously one meter apart into a very wide river flowing 0.5 m/sec (Fig. 30 – 24). The speed of the waves in water is 2 m/sec.

(a) Draw to scale the single crest spreading from each stone as seen by an observer on the shore at times 1 sec, 2 sec, . . . 5 sec. Join the points which mark the intersections of the crests.

(b) What would the line joining the double crests look like if the water were at rest?

16. Two ripplers 4 cm apart are generating waves at a frequency of 10 per sec in a tank where the wave velocity in still water is 30 cm/sec. Draw the nodal pattern if

(a) one rippler dips into the water 1/20 sec later than the other, and the water is at rest.

(b) the ripplers are in phase and the water flows at 10 cm/sec parallel to the line joining the ripplers.

17.* (a) In Fig. 30 – 10, where is the single slit, at A or A'?

(b) What is the significance of the other of these two points?

18.* Locate the points S_1, S_2, A, and A' (from Fig. 30 – 10) in the schematic diagram of the apparatus in Fig. 30 – 9.

19.* If there were no ether flow, the central maximum in Fig. 30 – 10 would lie on the perpendicular bisector of the line S_1, S_2.

(a) How would it be shifted if vacuum existed on both sides, but ether flowed only between the single and the double slit?

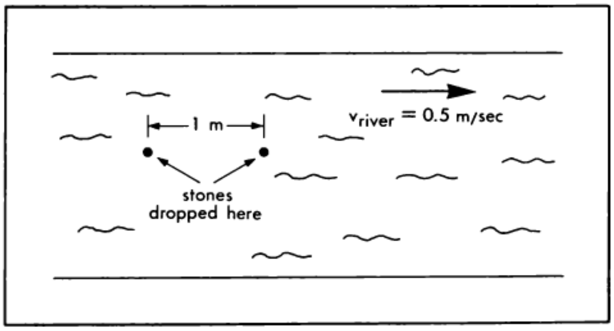
(b) How would it be shifted if vacuum existed on both sides, but ether flowed *only* beyond the double slit?

(c) How would it be shifted if vacuum existed and ether flowed on both sides?

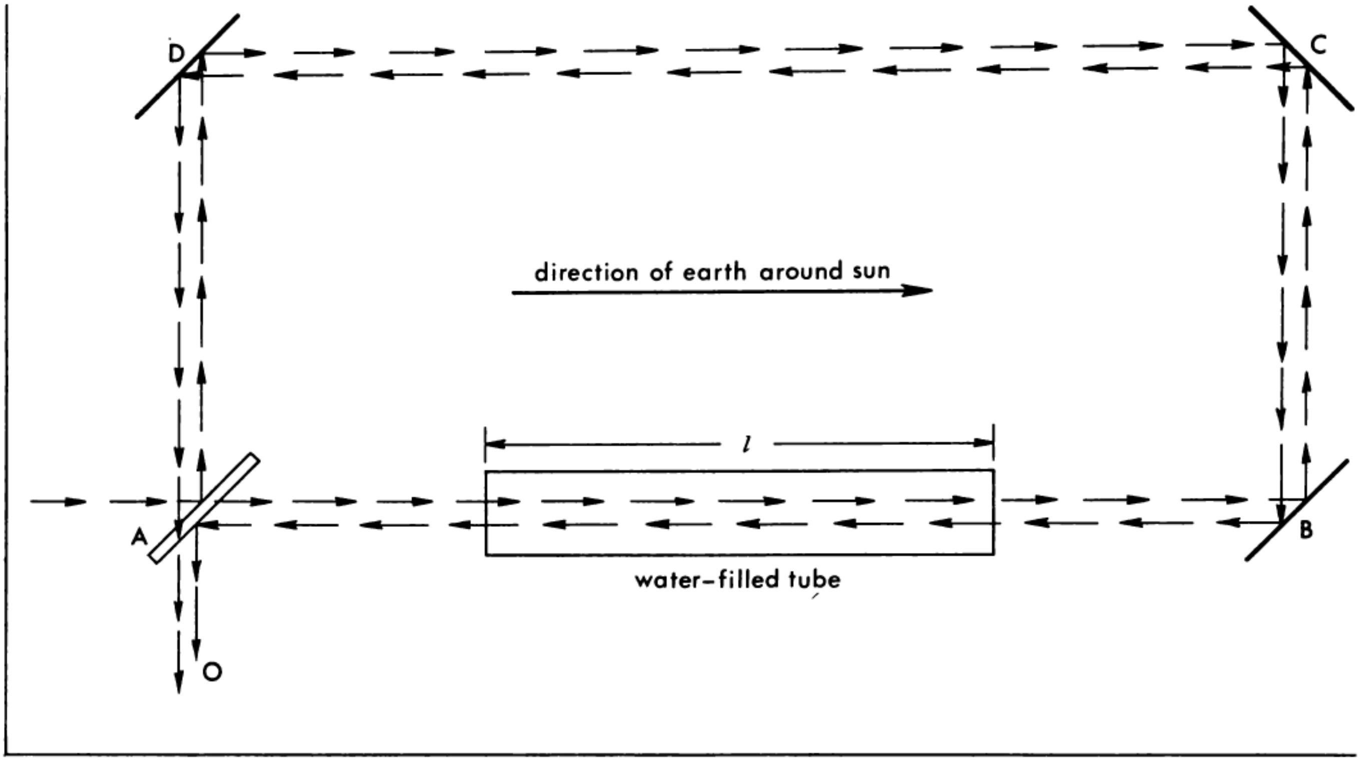
(d) How would it be shifted if vacuum existed beyond the double slit, but water was between the single and the double slit (and ether flow on both sides)?

20.* A parallel beam of light arrives at two slits at an angle α with the normal to the line joining the slits, and is diffracted. What is the phase difference between the light from the two slits if they are a distance d apart and the wavelength of the light is λ?

21.* In his original experiment, Fizeau also sent the light through his apparatus when the tubes were filled with flowing air and not with water. If the air flowed at 7.1 m/sec, what phase delay do you calculate?



30–24 For Question 15.



30–25 For Question 22.

22. As shown in Fig. 30 – 25, light from the source is split by the half-silvered mirror A and goes both clockwise and counterclockwise around A, B, C, and D. Between A and B there is a tube of water of length l. What is the difference, Δt, in the time it takes the two beams to make the circuit? Assume that the ether is at rest with respect to the sun, the tube of water points along the direction of motion of the earth around the sun, and velocities add by simple addition.

23.* What is the relativistic addition of two velocities $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ which are in opposite directions? (Section 30 – 10.)

24. Suppose the water in Fizeau's apparatus (Section 30 – 9) is moving at 30 m/sec. Calculate the shift in the fringes using (a) the simple addition of velocities and (b) the relativistic addition of velocities.

25. A spherical pulse of light spreads out from a point source.

(a) What is the difference in velocities of two parts of the pulse traveling in opposite directions as seen by a stationary observer at the source? (b) What is the relative velocity of one part with respect to the other part?

26. Could the expression $w = (u + v)\left(1 - \dfrac{uv}{(u + v)c}\right)$ be used for the relativistic addition of velocities? Check it for v or $u = c$, and for both u and v small compared to c. Does a change in sign of u or v make any difference?

27. Consider three galaxies, A, B, and C. Suppose that an observer in galaxy A sees B and C moving in opposite directions, each with a speed 0.7 c relative to A. (See Fig. 30 – 26). What is the speed of C as measured by an observer in galaxy B?

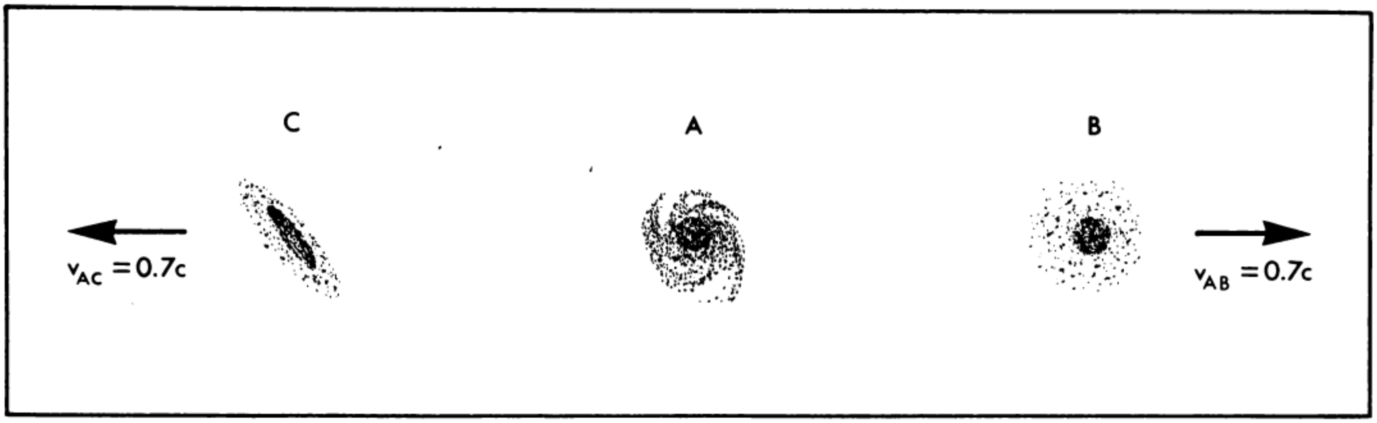
28.* According to the relativistic addition of velocities, what should Fizeau have seen when he substituted flowing air for flowing water in his apparatus? (See Problem 21.)

29.* Suppose that observer A in Fig. 30 – 14 presets his clock to T instead of zero. Observers A and B want to use a flash fired at A to synchronize their clocks. To what time should B preset his clock?

30.* An observer located a distance d to the *left* of A wants to synchronize his clock with that of A by using a flash given off at A. If A presets his clock to zero, how should this observer preset his?

31.* Distance AA' in Fig. 30 – 17 is vt as seen in the unprimed system.

(a) What is this distance as seen in the primed system?



30–26 For Question 27.

(b) Compute the difference between the two views of distance AA'. To what order in v/c do they differ?

32.* If $x' = 0$ in the expression $t = t' + vx'/c^2$, then $t = t'$. Which clock(s) in the unprimed system reads the same as which clock(s) in the primed system?

33.* Does the Lorentz transformation predict the correct expression for the relativistic addition of velocities, as the first-order transformations did in Section 30 – 12?

34. In Fig. 30 – 17, the clock of B' is a distance d from A', and is reading $t' = d/c$ as the flash arrives. To first order, observer C was placed at $x = d\,(1 + v/c)$ with his clock preset to $t = (d/c)$ $(1 + v/c)$.

What should the position and preset time have been for C, according to the Lorentz transformations?

35. Two relativistic trains are traveling on parallel tracks, the primed moving to the right relative to the unprimed at $v = 0.6\ c$. A flash goes off at O on the unprimed train just as O' on the primed train passes O. This flash is used by all observers on the unprimed train to synchronize their clocks with O. It is also used by all primed observers to synchronize their clocks with O'. Points O and O' are chosen as the origin for their respective reference frames. [Fig. 30 – 27 (a)].

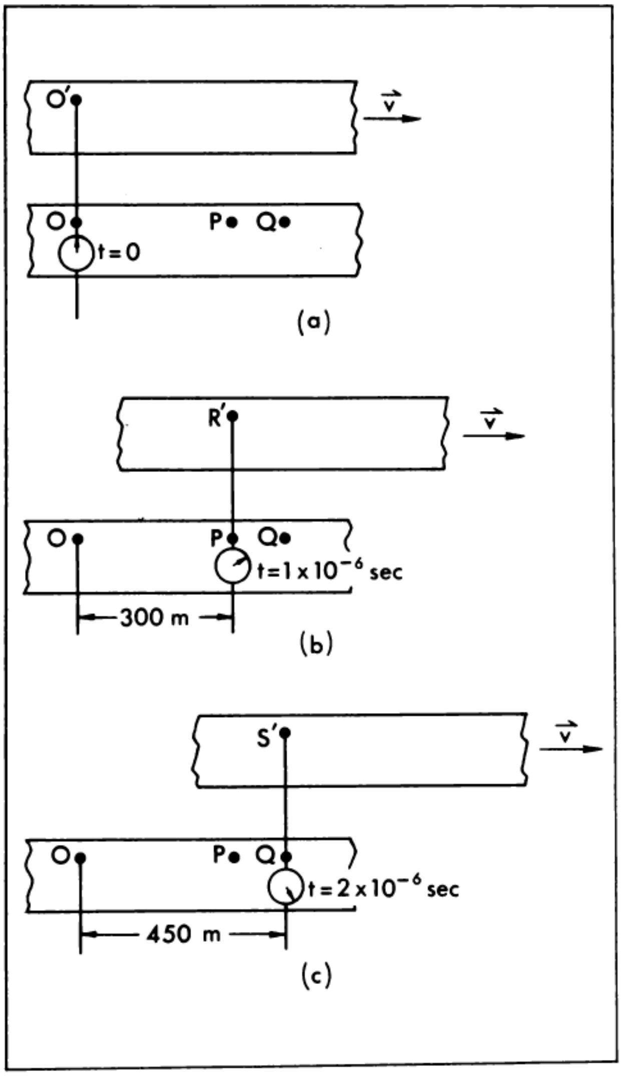
Point P is 300 meters to the right of O; Q is 450 meters to the right of O. When a point R' on the primed train is opposite P, the clock of P reads 1×10^{-6} sec, as in Fig. 30 – 27 (b). (Call this event 1.) When another point on the primed train S' is opposite Q, the clock of Q reads 2×10^{-6} sec, as in Fig. 30 – 27 (c). (Call this event 2.)

(a) What is the distance between the locations of the two events in each frame of reference?

(b) What is the elapsed time between events in each frame of reference?

36. (a) Is there another frame of reference (call it the double-primed system) in which the two events in Question 35 occur in the same place?

(b) What is the elapsed time between the two events in the double-primed system?



30–27 For Question 35.

37.* The synchronizing flash, triggered at A and A' in Fig. 30 – 15 arrives at B' and C at a later time (Fig. 30 – 17). Is this time interval a proper time in the unprimed system, in the primed system, or in neither?

38. If event A at a point x precedes event B at the same point as seen in one frame of reference, is the time order of the events preserved in every other frame of reference which moves with constant velocity with respect to the first one?

39. A transmitter on earth emits two pulses of light a time T apart. These are picked up by an astronaut moving away from the earth with speed v.

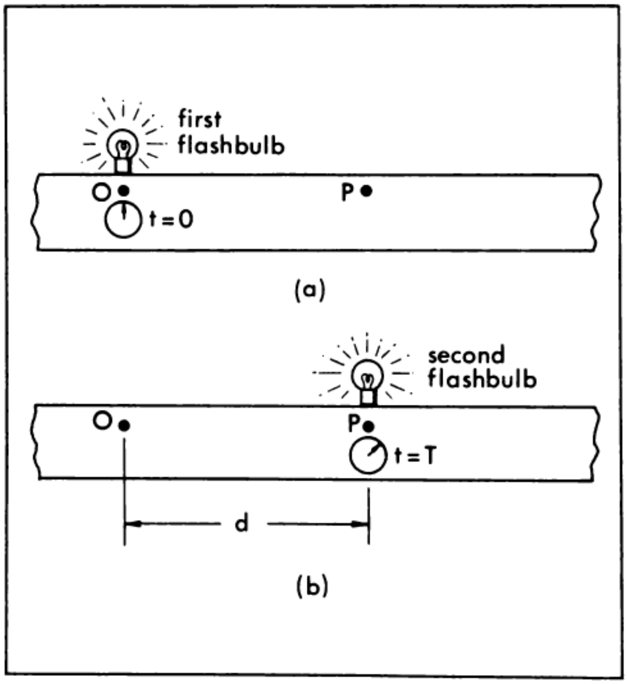
(a) Suppose the astronaut is a distance l from earth when the first signal is emitted. At this time all the clocks in the earth system read zero. What do these clocks read when the first signal is received by the astronaut? (Hint: imagine a clock at rest in the earth's frame but happily located adjacent to the astronaut as he receives the first signal.)

(b) What is the astronaut's distance from the earth when the second signal is emitted?

(c) What do the earth-system clocks read when he receives the second signal?

(d) What then is the time difference Δt between the arrival of the two signals as measured by observers at rest with respect to the earth? Is this a proper time?

(e) What is the time difference $\Delta t'$ according to a clock fixed to the space ship?



30–28 For Question 41.

(f) What is the wavelength of such signals of earth wavelength λ as detected by the astronaut?

(g) How fast must a star be moving away from the earth for the spectral line of 4000 Å to be seen on earth, as a line of 6000 Å?

40. An astronaut going away from the earth with speed of $0.2\,c$ is eating a meal and televises the procedure to the earth. On the table with the meal is a clock. A television observer on earth sees, on his TV screen, the astronaut's clock reading 12 noon when the meal is begun. When he sees this, the TV watcher sets his own clock to read 12. Some time later the astronaut is seen to finish the meal and the clock on the table is seen on TV to read 12:30 p.m. What does the TV observer's clock read when he sees this on his TV screen?

41. A flash which is used by all observers for clock synchronization goes off at O, the origin of the unprimed coordinate system, as in Fig. 30 – 28 (a).

A second flash goes off at P, which is a distance d to the right of O, at a time T later [Fig. 30 – 28 (b)]. Let us assume that this occurs after the arrival at P of the first flash; that is, $T > d/c$.

The two flashes thus occur in the unprimed system at different places a distance d apart; the time interval between the two events is T.

(a) At what speed must a primed coordinate system move relative to the unprimed system so that the two flashes occur in the same place in the primed coordinate system?

(b) What is the time interval between the two events in that primed system?

42. (a) At what speed must a primed system move so that the two events in Question 41 occur simultaneously in it?

(b) What is the distance between the two events in that coordinate system?

43. The distance from the earth to Sirius, assumed to be at rest with respect to the earth, is 8.8×10^{16} m (1 light-year $= 9.48 \times 10^{15}$ m). Suppose that at time $t = 0$, an astronomer sees an astronaut in a space ship passing the earth and going toward Sirius at a velocity $v = 0.995\,c$.

(a) How much older will the astronomer be when the astronaut passes Sirius?

(b) How much older will the astronomer be when he sees the astronaut passing Sirius?

(c) How much older does the astronomer calculate that the astronaut will be when the latter passes Sirius?

31 Photons

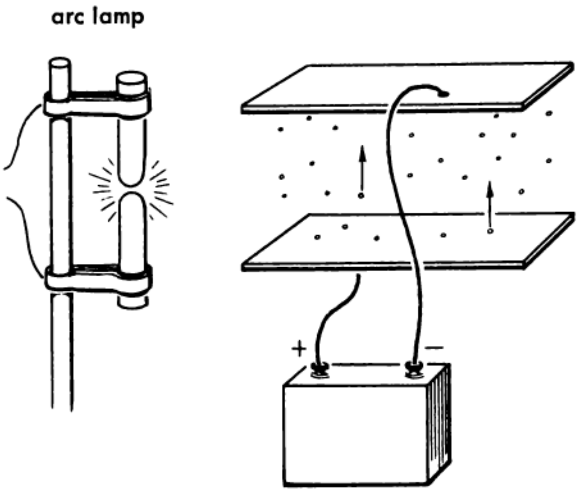
31–1 THE GRAININESS OF LIGHT

The wave theory has had convincing successes in describing all forms of electromagnetic radiation, from radio waves measured in kilometers down to the light waves which in our microscopes bring us a view of the smallest cell. We shall now give an account of experiments, some actually as they were done, some only illustrative of what might be done, in which the wave picture was subjected to a new kind of test.

That test was to see what light did to individual atoms. We cannot simply assume that light will behave in the same manner on an atomic scale as it does on a much larger scale. As a simple example, consider the experiments involving the refractive properties of light—light passing through a refractive medium is bent toward or away from the normal. If we use smaller and smaller blocks of refractive material, will Snell's law continue to hold? You know that when the thickness of the block becomes comparable to the wavelength of the light, diffraction effects become more important than the refraction. What happens if we go to a still smaller size—the size of an atom? Therefore, when we ask the question, "What does light do to individual atoms?" we must experiment further.

In the Millikan experiment, discussed for a very different purpose in Section 24 – 5, X rays—high-frequency light—expel individual electrons from tiny specks of matter. A good many tries were made of a related experiment: a kind of Millikan experiment in which X rays or ultraviolet light from an arc lamp falls upon large numbers of specks of material drifting between a pair of charged plates (Fig. 31 – 1). We ask what the wave picture of light predicts about the ejection of electrons from the tiny specks, and compare the prediction with what actually happens.

When light—that is, electromagnetic radiation—shines on a speck of matter, the electric field in the light wave pushes on the electrons in the atoms. If an electron is ejected from an atom, its ejection must be an effect of the electric field which forces it out of its normal motion



31–1 The light from the arc lamp illuminates a large number of droplets or specks of metal. These specks have been allowed to charge, and are floating between charged parallel plates; they are nothing but a large crowd of "Millikan oil drops." As the light strikes the specks, they may lose electrons. Each time an electron is lost, the speck will begin to drift upward toward the negative plate, for it is now positively charged by more than enough to balance its weight. Here and there, at random over the whole lighted region, a speck will suddenly start upward.

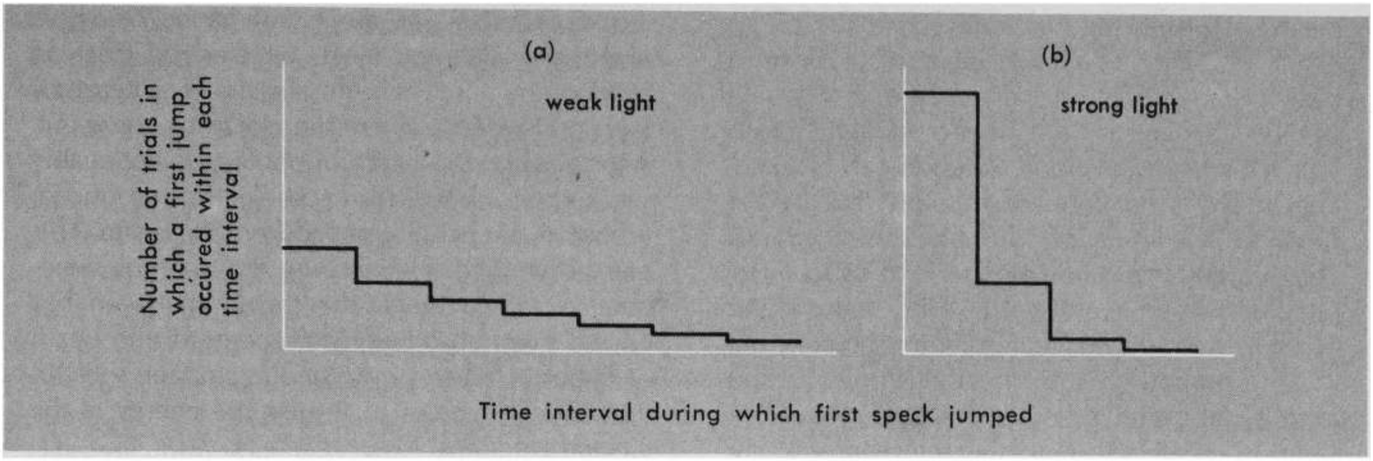
within the atom. (The magnetic field exerts a force perpendicular to the velocity of the electron and therefore does no work.)

Now, the electric field is rapidly changing in time, but at a given instant it has the same strength everywhere along the front of the electromagnetic wave. When such a wave hits an atom, therefore, it starts to shake the electrons. Since the wave looks the same over a large region of space, it shakes electrons in the little specks of matter in the same way at almost exactly the same time all over the region between the plates. If the light is weak, it will take the light wave considerable time to shake any electrons loose. Then, since there are many electrons in identical atoms pushed in the same way by the electric field of the light wave, a large number of electrons should be ejected at almost the same time. We thus expect to see, for rather weak light, a time during which nothing happens. Then, at one time all over the front of the light beam, individual specks should begin to lose electrons. With stronger light, the loss should come sooner; with weak light, later.

The actual result is astonishingly different. Instead of a definite wait, after which all the specks begin to lose electrons, it is found that a speck or two may jump immediately. These jumps show that the specks have already lost electrons and are now moving like the Millikan spheres under the influence of the electric field between the plates. The first specks to jump may be found anywhere in the beam. After them others will jump—one here, one there, with no particular pattern. After a while most of the specks will have lost electrons, but in a quite random way. Some may have lost several electrons; a few will have lost none at all.

If the light is made weaker, the waiting period is sometimes no longer. If the light is stronger, the wait is sometimes no shorter. We can make the specks as similar as can be, use our best identical spheres, and yet the individual specks behave differently (Fig. 31 – 2). Only the average behavior seems to be what we expect. On the average the specks will lose electrons as well in one part of a uniform light beam as in another. On the average, more electrons are driven out in one second by strong than by weak light; but this refers only to the average over many specks and many trials. In any particular case, how long it takes the first speck to respond seems to be a matter of chance. Once in a while, even the weakest light will cause the immediate loss of an electron; once in a while, even in the strongest beam, you will wait a fairly long time to see the first electron removed. The whole nature of this result seems to belie our well-founded view of the smooth oscillating wave front, everywhere uniform and moving in phase, displacing electrons in every speck by the cumulative action of the varying electric field.

The predictions of our picture of electromagnetic waves do not fit the facts, but we know of another simple model, long discarded, which we can now revive. Suppose that instead of the smooth plane wave front, the beam of light is represented by a collection of fast-moving tiny particles, like a volley of bullets or an air-driven blast of sand (Fig. 31 – 3). In such a beam, a speck would be hit squarely by a single particle of light only by chance. The hits would appear now here, now there, over the whole of the beam. The moment when the first speck was hit would also be a matter of chance. Even if the beam was weak—with only a few particles

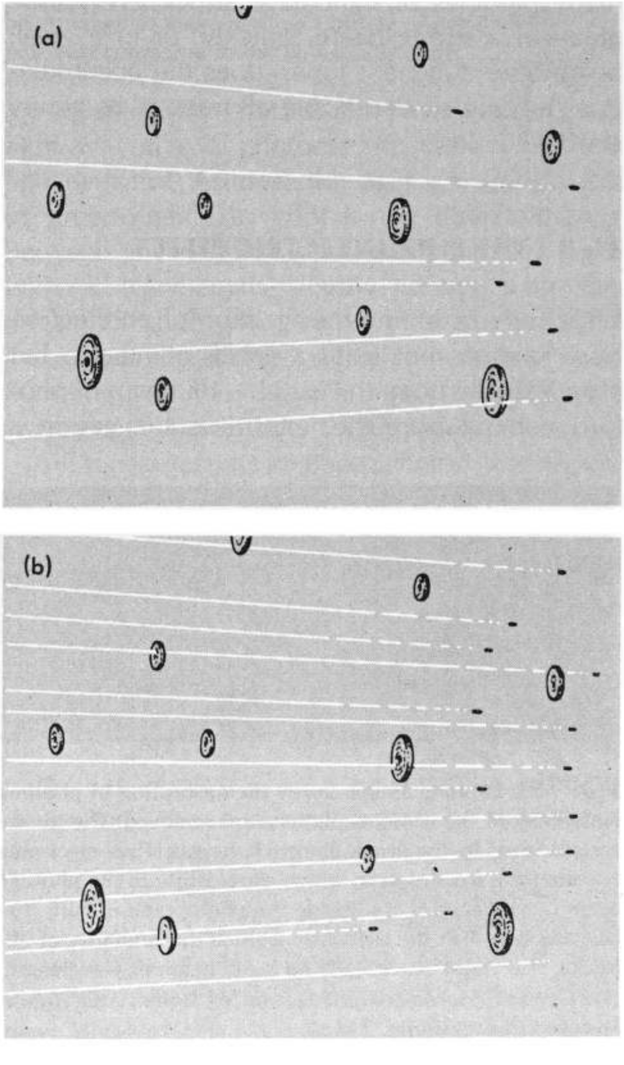


31–2 **In the apparatus of Fig. 31–1, we turn the light on suddenly and watch for the first speck to start upward. This time is then recorded. The graphs show the results of many such time-lag measurements. In (a) the experiment was done with weak light; in (b) with much stronger light. Everything else remained unchanged. The total number of trials in each group of experiments was the same. On the average, the strong light ejects electrons more quickly than the weak light, but even in weak light some electrons are ejected with no appreciable lag. Experiments of a more advanced sort have pushed this minimum lag time below 10^{-8} second.**

crossing a given area in a second—it might happen that a particle of light strikes one speck immediately after the beam is turned on. On the other hand, in a strong beam, the dense rain of particles might fail to make any direct hit for a noticeable time. Only on the average would it be true that in a strong beam the first hit would come soon, and the others follow in rapid sequence. This picture gives an accurate and convincing description of the way light really ejects electrons from the little specks of matter.

On this picture light comes in grains. It arrives in little bundles which we shall call *photons*. The photons are the bullets or particles of light; and when they hit a speck of matter, they may eject an electron.

Our experiment directly detecting the ejection of electrons by a light beam is not the only



31–3 **The close analogy between bullets fired in a random volley at a set of targets, and a light beam ejecting electrons. If the volley comes from only a few guns (a), the "beam" of bullets is weak. Most of the time, you would wait a long time to see a direct hit. But occasionally a lucky hit would immediately follow the start of firing. With many rifles firing (b), the "beam" is intense. Hits become frequent, and on the average you wait less time for the first one. But the first hit might come no sooner, or even later, than an early first hit in a weak beam. Just which target would be hit next, of course, is a matter of chance.**

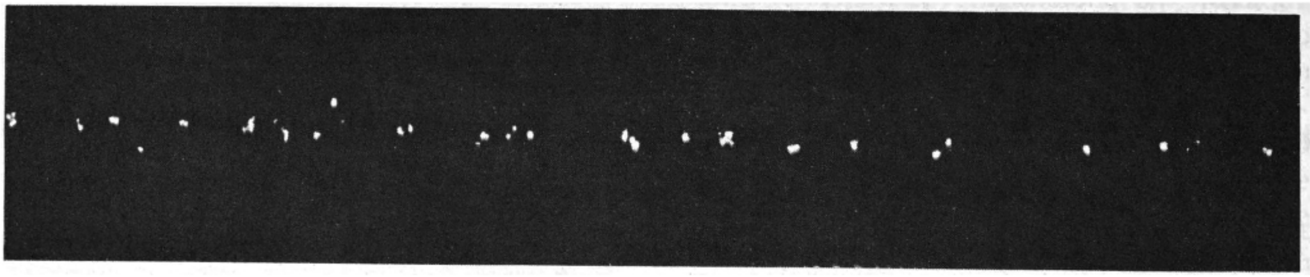
means of displaying the grainy character of the beam of light. A photographic plate bears a layer of very small crystals, each one of which can be changed by the impact of light into a form which the developer solution can blacken. The image is not formed smoothly, but by the random accumulation of blackened grains. This process is too complex to lead us to interpret light from it alone—the film grains differ in their atomic constitution, for example—but it adds support to the direct electron-ejection experiment. With other devices using ejection—the photoelectric cell, the television-camera image tube—one can watch an image being formed point by point by light. In very weak light, the successive "hits" are so slow that the eye can easily follow them. The image, whatever it may be, is painted slowly by the accumulation of random hits, which gradually pile up, many wherever much light falls, few wherever the illumination is weak. Only when very many particles of light have struck every little area—in a strong beam, quickly, or in a weak beam after sufficient time—does the accumulation become so dense that all trace of its grainy structure is lost.

31–2 THE PHOTOELECTRIC EFFECT

Let us now examine the ejection of an electron by a photon, not from a speck of matter, but from a single atom. In Fig. 31 – 4 a beam of photons coming from the left strikes the gas in a cloud chamber. As the beam penetrates, here and there a photon ejects an electron from an atom of the gas. The little clusters of waterdrops form at the place where the electrons come out. After a while the beam of photons is noticeably weakened; indeed, for each electron produced, a photon has been removed from the beam. This can be verified by watching the photons penetrate so far into the gas that the original beam has all but been absorbed. The energy of the beam of light has passed into the gas. In fact, most of that energy appears as the kinetic energy of the ejected electrons.

Each electron makes a track of about equal length wherever it may start. (There are no very long tracks in Fig. 31 – 4; they are all about equally short and stubby.) So each of the emitted electrons must have had almost the same energy. From this point of view all the events are the same. And in particular each photon that causes an event is the same. A weaker beam means fewer photons, rather than photons of different energy. When we study the energy balance more closely we find that the energy is conserved in each individual event, and we again check that a weaker beam only means fewer photons.

Such a cloud-chamber study cannot be made with photons of visible light. In the experiment shown in Fig. 31 – 4, X rays were used, forming a beam of electromagnetic radiation of wavelength only a few tenths of an angstrom. At this wavelength the individual process is then made obvious. With X rays of shorter wave-



31–4 The cloud chamber shows the absorption of photons, one by one, and the weakening of the beam as photons are removed. The photons enter from the left, and are absorbed by the argon atoms of the gas. They eject photoelectrons, whose tracks acquire little fog droplets, which show white in the photograph. You can see that the beam is weakened if you divide the photograph in half; there are 17 tracks in the left half and only 9 in the right half. Notice the similarity of all the short photoelectron tracks. The "light" is actually an X-ray beam of wavelength about 0.2 angstrom. (W. Gentner, H. Maier-Leibnitz, and W. Bothe, Atlas Typischer Nebelkammerbilder, Springer Verlag, Berlin, 1940.)

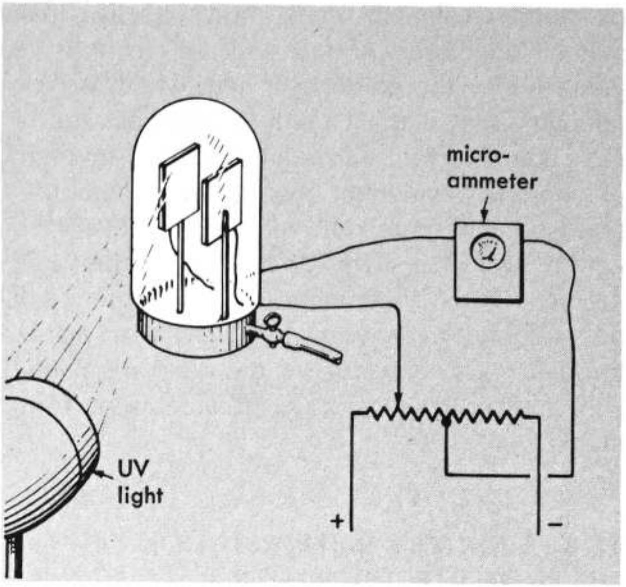
length the photoelectrons are knocked out harder and the identical nature of the individual events is even easier to see.

The use of a beam of ultraviolet light allows a more detailed, if a less direct, study of the process of ejection of an electron by light. This process is called the photoelectric effect. It can occur at a metallic surface, in a liquid, or in the individual atoms of a gas.

To study the photoelectric effect, we can use a simple device (though not an easy one to build). It is a glass bulb, evacuated to a first-class vacuum, with two small and really clean copper plates mounted inside (Fig. 31 – 5), and two leads for electrical connections to the plates. We apply a potential difference of several volts between the plates, and shine light on the inner surface of one plate. If we illuminate the positive plate, there is no measurable current within the vacuum. But if instead we illuminate the negative plate, there will be a current. Evidently electrons are flowing across the vacuum. If the light is cut off, the current ceases. If the beam of light is turned on again, the current begins without delay – direct measurement has shown that the lag in the photoelectron emission is certainly shorter than 10^{-8} second. The current is not large, but it is within easily measurable range; a copper plate a few square centimeters in area will yield microamperes of photoelectric current in direct sunlight. The current will increase and decrease in proportion to the intensity of light over a very wide range. Similar tubes are widely used in practical applications, from opening doors to reading the sound from movie film.

Our simple phototube confirms what we knew: when light strikes, electrons are released from the surface of a substance. The stronger the light, the greater the number of electrons ejected. The number of electrons ejected is proportional to the number of photons. All this makes sense, but we are in for a surprise when we change the wavelength of the light used. The copper photocell will give no current in even the strongest beam of red or of green light; only ultraviolet light, or light of still shorter wavelength, will work.

Perhaps this effect occurs only in copper. Let us try another substance. We continue the experiment with a tube in which the plate surfaces are coated with potassium. Now visible blue light will give a photoelectric current but red light will not. A threshold, as it is called, exists in green light. (In copper the threshold was between the visible and the ultraviolet.) Light with a wavelength slightly longer than the threshold value will give no photocurrent even if it floods the surface, while some current will flow in even the weakest beam of light with wavelength shorter than the threshold value.

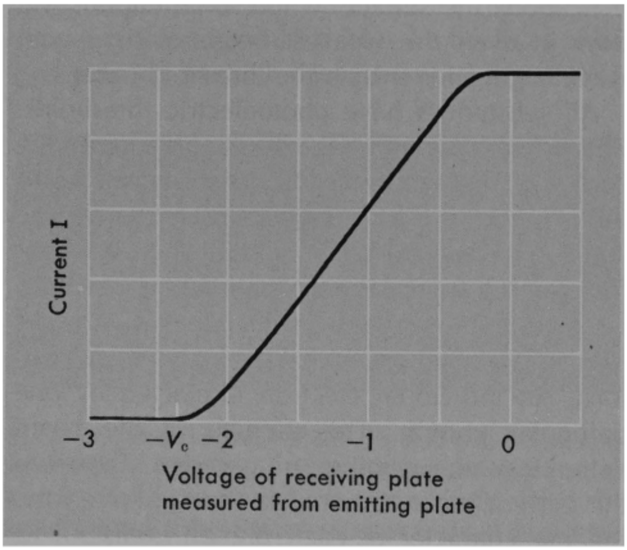


31–5 The photoelectric cell. Two copper electrodes are mounted in a well-evacuated glass bulb, and leads are brought out as shown. The circuit allows for adjustment of the potential difference between the plates; the moving contact is connected to the unilluminated plate, and can be arranged to make it either positive or negative in potential compared to the illuminated plate.

All substances have photoelectric thresholds. The wavelength of the threshold depends on the nature of the substance. But there is one common point: for long waves the photoelectric effect does not take place, and for short waves—for light of high enough frequency—it does. No picture of a wave will suffice to explain these results, but they fit well with the idea of photons. An individual electron is ejected by one particular photon. The strength of the beam determines the number of photons. Therefore the current is proportional to the light intensity. But the character of each photon collision, its ability to eject an electron, does not depend at all on how many others are present. It is a property of every single photon, and must somehow depend on wavelength of the light. Light of short wavelength seems more effective than that of long. Every surface will let electrons go when hit by the photons of light which is blue enough; a few surfaces will let electrons go for the photons of red light—and those of any light of shorter wavelength. Some artful combinations, notably the alkali metal cesium dissolved with oxygen in a silver surface, have a threshold in the infrared; but metals like platinum will not respond to any of the light that passes through glass; only use of the deep ultraviolet will suffice to generate a photocurrent from them.

31–3 EINSTEIN'S INTERPRETATION OF THE PHOTOELECTRIC EFFECT

We have seen a little way into the mechanism of the photoelectric effect. We know that it involves some detail of the photoelectric surface. We also know that it always requires light of high rather than low frequency. This suggests that the photons differ in some way dependent on the frequency. That photons of red light are different from those of blue light is only reasonable. Photons must somehow carry the information which shows up as the wavelength when light waves interfere. To find out in what way the frequency is contained in the photons we must do something new. We must find out how violently the photoelectrons are ejected.



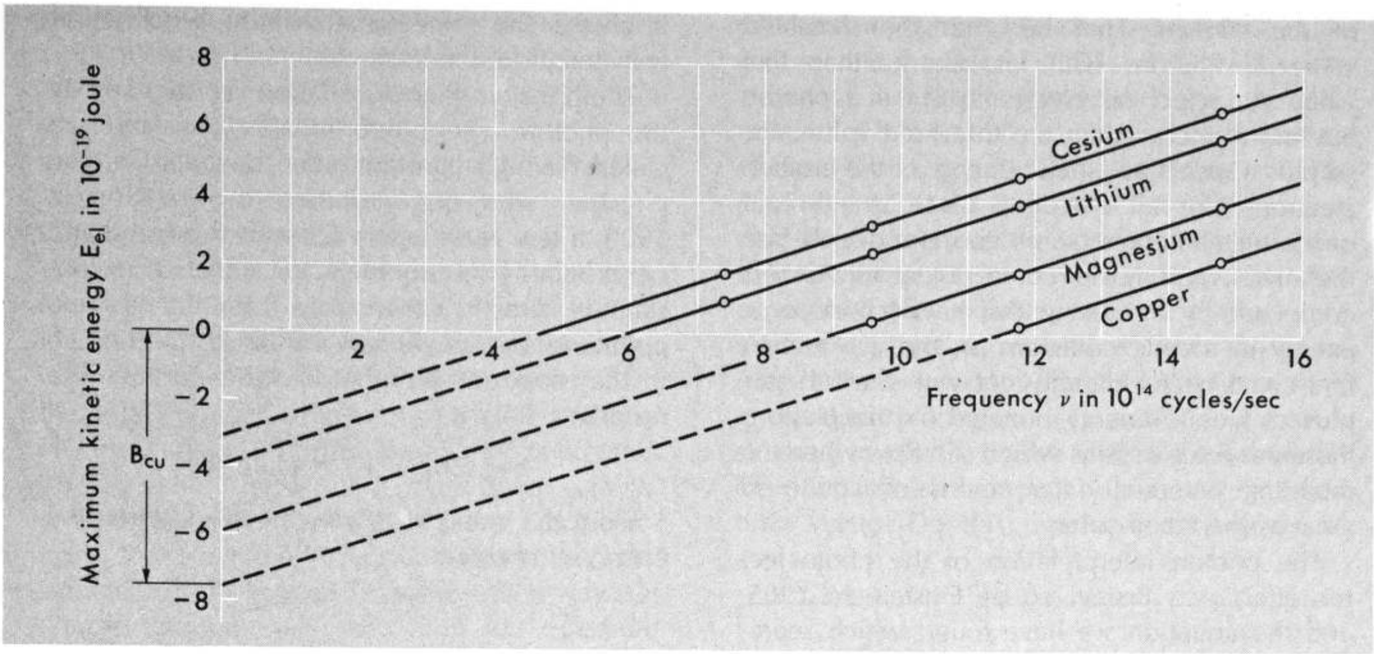
31–6 Current of photoelectrons vs. potential difference between the plates for a given kind of light. Note that the current does not stop until the receiving plate is a volt or so negative. This value is the cutoff potential V_c.

In the apparatus of Fig. 31 – 5 we can adjust the magnitude of the applied potential. We get the maximum current for any particular choice of surface and any given illumination by making the receiving plate slightly positive with respect to the emitting one. No increase in current follows upon increasing the voltage; evidently we are picking up all the photoelectrons ejected (Fig. 31 – 6). Now we can turn the voltage down and reverse the potential, making the emitting plate the positive one. Then the electrons are attracted back to the plate as they seek to leave it; but the photocurrent does not vanish. Apparently the ejection of photoelectrons by light imparts to some of the electrons enough energy to overcome the retarding electric potential difference between the plates. Finally, as the retarding potential is increased, at some particular retarding potential difference, the current stops. Even stronger light (if the wavelengths remain unchanged) will not restore any photocurrent. None of the photoelectrons have enough energy to cross between the plates against this retarding potential difference. We call it the cutoff potential. (See Fig. 31 – 6.)

The cutoff potential V_c is a means of studying the energy imparted in each single act of electron ejection. The cutoff potential difference times the one elementary charge of each electron tells us the energy that must be taken from the fastest electrons to stop them from going all the way across from one plate to the other. It therefore tells us the maximum kinetic energy E_{el} with which these photoelectrons are emitted:

$$E_{el} = (1 \text{ elementary charge}) \times V_c$$

With the cutoff potential as a tool, we can measure the maximum kinetic energy of the photoelectrons ejected in a series of experiments on light beams of different colors. When we are at the cutoff potential difference with one wavelength, we find that we can always restore some photocurrent by using light of shorter wavelength. This observation tells us that the shorter the wavelength of the light we use, the more energy the photons can give to the ejected electrons. We can measure just how much more by finding the cutoff potentials at



31–7 The results of experiments with the apparatus of Fig. 31–5, augmented by the use of various light sources and filters of different color. The maximum kinetic energy E_{el} is plotted against the frequency ν of the light used.

shorter and shorter wavelengths.

If we plot the results of a series of experiments with light beams of different colors striking the surfaces of different metals, a simple and fundamental relation emerges. It is shown in Fig. 31 – 7, where we plot the maximum kinetic energy of the photoelectrons against the frequency ν (rather than the wavelength) of the light.* For any particular substance, the curve of maximum kinetic energy versus frequency is a straight line with a slope which we shall call h. And this slope h is the same for all materials. Only the intercept, B, changes from one material to another. The graphs for a few typical metals are shown in Fig. 31 – 7. Consequently, the graphs for all substances can be described by the relation

$$E_{el} = (1 \text{ elem. charge}) \times V_c = -B + h\nu$$

where h is always the same, B is constant in any particular experiment, and ν is the frequency of the light which changes with the spectral color.

Here we see that the experimental information clearly distinguishes between the role of the light and the role of the photoelectric surface. The surface is represented solely by the term $-B$, which is the intercept of the photoelectric curve indicated on the left in Fig. 31 – 7. This term is independent of the light, for we can change the frequency ν as much as we like without influencing it. Consequently, it only contains information about the nature of the surface. On the other hand, the light is described by the term $h\nu$, which is always the same for light of the same frequency. (Because the slope h is the same for all substances, this term has nothing to do with the nature of the photoelectric surface.) This term tells us in what way the photon carries the information about the frequency of light. Each photon is able to give more energy to an emitted photoelectron in direct proportion to the frequency; that is, E_{el} increases with ν.

On the photon picture we can interpret the last equation completely. Each photon carries (or perhaps the photon *is*) the energy $h\nu$. When the photon hits the surface it disappears, and the energy $h\nu$ is transferred. Some of this energy is needed to bring the photoelectron out of the surface. That becomes the threshold value B. Photons with less energy than that value can eject no electrons. But if a photon has more energy than the threshold value, the surplus appears as kinetic energy of the emitted electron. Not all electrons come straight out of the metal surface; some go more deeply into the metal, and never come out at all. Others come out by indirect paths, having lost some energy in atomic collisions on the way out. A few come nearly directly out with the full surplus of kinetic energy donated by the photon; these are the electrons which still flow when the retarding potential difference is not quite so great as the cutoff value.

---

*In modern physics the frequency is almost always represented by the Greek letter ν (read nu) rather than by the letter f. We shall follow this standard practice throughout the rest of the book. When you read modern physics in other books, the symbols will then be familiar to you. For example, the relation between the speed of light, its frequency and its wavelength is written $c = \nu\lambda$, meaning just the same thing as $c = f\lambda$, which you used in Part II.

The photon interpretation of the photoelectric effect was first given by Einstein in 1905, and the equation we have found which represents the conservation of energy in the photoelectric process is known as Einstein's photoelectric equation. But the experimental evidence that we used to establish this relation was not available to Einstein. He invented the whole picture on the basis of far less evidence, and predicted the experimental results which we have used. In this work Einstein built upon a result obtained by Planck. In 1900 Planck had explained the spectrum of light inside an insulated enclosure at constant temperature.* His explanation implied that light was emitted and absorbed in energy bundles $h\nu$ in size. Einstein now pointed out a number of pieces of evidence for the reality of these bundles. He made them into particles of light, into photons; and at the same time he used them to found the theory of photochemistry and forecast the results of later photoelectric experiments.

The photoelectric effect has been explained by the photon picture. In any beam of light of frequency ν, each photon carries an energy

$$E = h\nu = \frac{hc}{\lambda}$$

This is the famous Einstein-Planck relation. A beam of light is more or less intense according as it has more or fewer photons per unit time; but the energy per photon, the quantum energy, is always the same for a definite frequency or wavelength of the light.

The Einstein-Planck relation is the key to the photon. The proportionality constant h is called Planck's constant, after the great German physicist who first proposed this relation in 1900, a few years before Einstein showed that it would clarify the photoelectric effect. This constant is directly measurable from the last experiment; for, as we saw in the last section, h is the slope of all photoelectric curves. The result is

$$h = 6.62 \times 10^{-34} \text{ joule-sec}$$

From the value of h, we can quickly find the energy of photons.

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6.62 \times 10^{-34} \text{ joule-sec})(3.0 \times 10^{8} \frac{\text{m}}{\text{sec}})}{\lambda} \\ &= \frac{1.99 \times 10^{-25} \text{ joule-meter}}{\lambda} \end{aligned}$$

Also, there are 10^{10} angstroms/meter and 6.25×10^{18} electron volts/joule. Therefore,

$$\begin{aligned} E &= (1.99 \times 10^{-25} \text{ joule-meter})(10^{10} \text{ Å/m}) \times \frac{6.25 \times 10^{18} \text{ ev/joule}}{\lambda} \\ &= \frac{1.24 \times 10^{4} \text{ ev-angstroms}}{\lambda} \end{aligned}$$

Visible light has a wavelength of about 5000 Å. Consequently, an average visible photon carries the energy

$$\begin{aligned} E &\approx \frac{1.24 \times 10^{4} \text{ ev-Å}}{5000 \text{ Å}} \\ &= 2.5 \text{ electron volts} \end{aligned}$$

The whole range of visible-light photons extends from an energy of around two electron volts for the red to somewhat above three for the blue.

These energies, of a few electron volts per photon, are exactly in the lowest energy range which can cause important chemical changes, presumably—as we shall see in the next chapter—because they are energies high enough to disturb the electrons within an atom or even to eject them. For these energies, then, but not for lower values, photographic chemicals, photoelectric surfaces, and the retina of the eye itself are sensitive. The differences among the "visible" ranges for such devices are exciting, but they are really rather small. No known photosurface will respond to infrared light which is even ten times longer in wavelength than the limiting response of the human eye.

---

*An insulated enclosure with a small hole in one wall is commonly called a "black body," because it absorbs radiation almost perfectly through the hole. Previous theories about the frequency distribution of radiant energy in a black body had been in violent disagreement with the experimental facts.

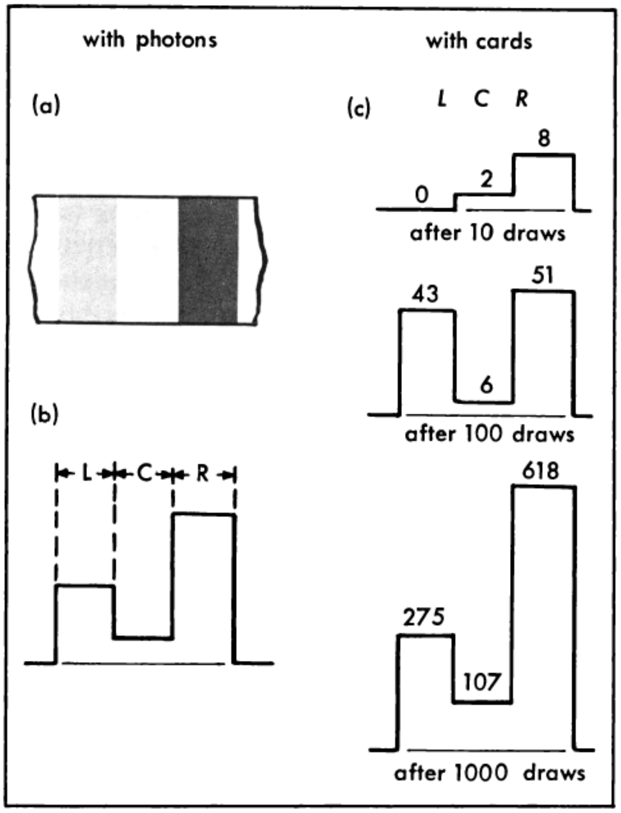
From the value for the energy of a single photon, we can find the number of photons streaming by in any radiation beam whose total energy rate we know. Then it is interesting to compare the total number of photons striking a photoelectric surface with the number of photoelectrons ejected. Even for a very good surface, only a few percent of the photons eject electrons. The rest send electrons deeper into the surface, or are wasted in some other way. Of course, photons with less than the threshold energy cannot eject electrons at all.

31–4 THE ORDERLINESS OF CHANCE

The ejection of individual electrons from a piece of metal is quite random. Some electrons are emitted immediately after the illumination, while some are delayed by various, unpredictable times. But the overall picture of electron emission is certainly one of orderliness. On the average, the specks will lose the same number of electrons in one part of a uniform light beam as in another. Also, on the average, more electrons are driven out in one second by strong light than by weak light.

In Chapter 21 you saw that a regular pattern is built up from a succession of equally probable events. It is easy to convince yourself that the accumulation of individual events, all going on by chance but some more probable than others, will also form in the end a regular and smooth pattern. There is no substitute for actually performing an experiment in which you watch meaning emerge from what appears to be a sequence of wholly meaningless and sporadic events. It is not practical to simulate the myriad grains of light which form a good photographic image, but we can consider a crude pattern formed by a smaller number instead.

Suppose the bombardment of some receiving plate by a light beam results in the bar pattern of Fig. 31 – 8 (a). More of the "grains" of light must in the end fall on the regions of the photographic plate that become dark than on those that remain lighter. We can represent this pattern by the graph of Fig. 31 – 8 (b), where the height of the graph is proportional to the total number of hits on that same region in the picture of Fig. 31 – 8 (a). Now we prepare a set of cards, marked *L*, *C*, and *R* for the left-, center, and right-hand bars of the pattern. We prepare cards in a number proportional to the total number of hits expected; say, 3 marked *L*, 1 marked *C*, and 6 marked *R*. This fairly expresses the assumed probability of a hit in each of the three places we distinguish. Now draw a single card blindly from a well-shuffled mixture. Read what it is, then put it back, reshuffle, and draw again. Do this over and over again. At



31–8 The growth of pattern out of chance. In (a) we see a bar pattern of light and dark, representing some sort of interference fringes. In (b) the same pattern has been graphed, showing how the light intensity varies with position. In (c) we see plots of an analogous experiment in which a similar pattern is built up by random chance draws from a collection of cards whose probability of draw has been correctly adjusted. The patterns made with more and more individual draws come closer and closer to the expected pattern, beginning with an apparently disorderly arrangement showing no resemblance to the final result. The ordinary light source presents us images which are painted out by some 10^{12} or even many more photons per second.

each draw, note the letter the card bears, and plot the results of each draw by adding to the height of the appropriate vertical bar on a graph constructed like Fig. 31 – 8 (b). At first, no clear sign of the pattern will be present, except very rarely by a remarkable accident. Later on, the pattern will begin to emerge, and if you are patient enough to draw several thousand times, the pattern thus randomly built up is almost sure to be as regular as any graph you could deliberately construct. The results of one such experiment are plotted in Fig. 31 – 8 (c), and your own experiment should give you a still better idea of how such patterns can arise. Of course, the grains of light are not cards; we are not really experimenting with light, but with the laws of chance and probability. But many experiments have demonstrated that the formation of images by light follows in quantitative detail exactly the laws obeyed by the random draw of shuffled cards.

With this model we produced the final regular pattern and also the irregular fluctuating build-up of the pattern that can actually be observed.

It is just for this reason that the finely detailed images caught by a TV camera or a photographic camera can be formed only in sufficient light. The inviolable minimum of light needed is set by the graininess of light. No matter how sensitive the receivers of the light may be, the image cannot be made unless there are enough of the light packets to form it accurately. It is striking that the effects of the intrinsic graininess of light have been demonstrated to occur for the eye just about at the dimmest light which a man can see. Our eyes are developed about as far as any light receivers working with the same size of lens, the same time of response, and the same band of wavelengths can possibly be. They have reached very close to the natural limit.

31–5 GRAININESS AND INTERFERENCE

In weak light we find the individual effects of grains. Little bundles of light come along at random, now here and now there, arriving in greater number where the intensity is higher and in smaller number where the light is weaker. This grainy picture of light is far different from our usual view of waves. It seems to have nothing to do with interference and diffraction. Can it be that the wave properties demonstrated so clearly in interference experiments apply only to strong light, where somehow the many individual "grains" of light interact one with another to produce the wavelike results? Or do the patterns built up by weak light also show interference? This question was directly tested fifty years ago by Geoffrey Taylor, a young student at the University of Cambridge, who later became Sir Geoffrey for his lifetime of successful work in physics.

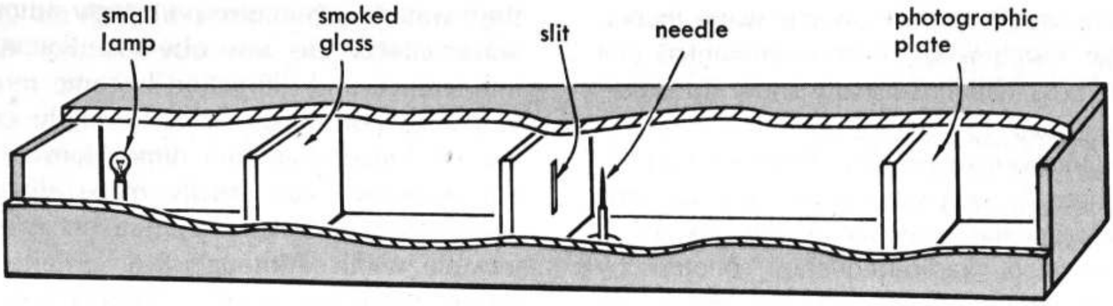
In a lighttight box Taylor set up a small lamp which cast the shadow of a needle on a photographic plate (Fig. 31 – 9). He chose the dimensions so that the diffraction bands around the shadow of the needle were plainly visible. Then he reduced the intensity of the light. Longer and longer exposures were needed to get a well-exposed plate. Finally, he made an exposure lasting for several months (which, the story goes, he spent sailing). Even on this plate the diffraction bands were perfectly clear, though he could show for this very faint light that only one photon—one of the packets of light—was in his box at a time. Interference therefore takes place even for single particles of light. It works for every single photon.*

Taylor's experiment is not the only experiment that points to this conclusion. There are many. Nowadays a photocell working an amplifier can make the arrival of the photons audible. It is interesting to hear the clicks which signal the painting out by individual photons of a two-slit diffraction pattern. While you watch and listen, the pattern is formed by photons coming in one by one. All interference experiments work this way: the photons come one at a time. So we are forced by direct experimental evidence to believe that wave behavior applies to individual photons.

What do we mean by the wave behavior of an individual photon? We shall have more to say on this by no means simple point. For the mo-

---

*Taylor knew that there was only one (and often no) photon in his box at one time because from the energy of one photon and the intensity of the light he could find the number of photons in any given length of light beam. He found the photons must be spaced farther apart on the average than the length of the box. So usually there was not even one photon in the box. You can do this kind of calculation after studying the next section.



31–9 **G. I. Taylor's experiment, proving that the interference pattern was correctly given even if only one photon was present in the apparatus at a time. He found good, sharp fringes in the shadow of a needle cast by light from a slit, even when that light came from a dim gas flame through such absorbers that an exposure of three months was needed.**

ment, we shall be satisfied to say that the electromagnetic wave must somehow fix the probability of appearance of a photon. Where the field in the wave is strong, there we shall very probably find photons; where the field is weak, photons will occur with small probability; where the field vanishes, no photons occur at all. When it interacts with the photographic emulsion or the photoelectric cell, no single photon is spread out over a whole interference pattern. The photon hits at a definite place. But for each photon, the probability is highest that it will arrive where the interference pattern is most intense. The probability of arrival at a region of half that intensity is half as big, and so on. The probabilities of arrival for each photon have just the pattern of the intensity of the interference fringes (the bright and dark interference bars). So when many photons have arrived, the interference picture is painted out by their cumulative effect.

31–6 PHOTONS AND ELECTROMAGNETIC WAVES

Where is the wave picture in all this? It may seem to have no relationship to what we have just been discussing, but we cannot forget all the evidence for light waves nor the evidence for electromagnetic waves of all kinds. Somehow photons and electromagnetic waves must fit together, and we already noticed that this is possible for average effects involving great numbers of photons. In Section 30 – 5 we saw that interference patterns—a typical wave effect—are painted by photons landing one at a time at many different points. Here we shall look again at the relation between waves and photons.

First we must establish that photons are compatible with all we know about electromagnetic waves in the region of wavelengths from millimeters to miles. We make and detect these waves, not by using single electrons or atoms, but with radio tubes and antennas, through which charges flow back and forth. And we have plenty of evidence that the wave picture agrees with the facts of radio transmission.

In that range the frequencies are small, relative to atomic values. A frequency of a few kilocycles represents the low frequency limit of radio technique, while a frequency of a million megacycles is beyond the most advanced radar and microwave techniques, and lies at the margin of the deep infrared. Yet even these highest radio frequencies, with the tiny atomic constant h, give a photon energy which is really very small.

At one megacycle, for example, the photons in a radio wave have the energy

$$\begin{aligned} h\nu &= (6.62 \times 10^{-34} \text{ joule-sec})(10^{6} \text{ cycles/sec}) \\ &= 6.62 \times 10^{-28} \text{ joule} \end{aligned}$$

Because there are 6.25×10^{18} electron volts per joule, this is

$$h\nu \times 4 \times 10^{-9} \text{ electron volt}$$

Such small photon energies mean that a radio signal cannot be detected unless it contains very many photons. The lower limit of a good communications receiver is about 10^{10} photons/second. Among so many photons the average behavior alone is detectable. The result is the smooth transfer of energy from wave to matter which we would expect on the wave theory. The large numbers of photons guarantee that the graininess will not usually show up unless we try very hard to detect it.

We might even suspect that there are no photons in light of long wavelength. But we now have experimental evidence of the absorption and emission of "radio-frequency" photons by atoms and molecules (see Chapter 33). So we know that both the photon and the wave description apply.

As we go to higher frequencies, we pass to the region of visible light. Here the number of photons is large in any light beam that appears bright to our eye. Sunlight, for example, represents a rain of some 10^{17} photons per square centimeter per second. In such a burst the individual arrivals of photons are not easily noticed. At the other extreme, however, we can still make out the glow of a faint luminous source which sends to our dark-adapted eye a hundred photons on the area of the pupil each second. So for visible light, both the wave and the photon sort of observation are possible.

The photographic plate and the photoelectric cell are, like the eye, capable of responding to the smooth flow of intense or moderate beams no less than to the fluctuating and chancy arrival of a few photons every second. It is for this reason that it was with the study of visible light, and not in the study of the long-wavelength portions of the electromagnetic wave spectrum, that the photon idea was first conceived and found satisfactory.

As the frequency of light increases and the photon energy rises, the photon description appears more and more dominant. The photon energy in the X-ray region is large enough so that single photons are easily detectable. The arrival of a single gamma-ray photon—associated with still higher frequency and correspondingly shorter wavelength—is energetically a sizable event. It can be noted with good probability in a whole variety of detectors. Gamma-ray photons, which have energies as great as 10^{14} electron volts, have been found in the high atmosphere. There they are produced by cosmic-ray particles reacting with the atoms of the air. Such a photon produces tens of millions of recoil electrons.

On the other hand, as the frequency rises and the wavelength correspondingly diminishes, wave effects are less obvious. For example, interference and diffraction become more difficult to demonstrate. The wavelengths of X rays are no longer than the dimensions of atoms. For them we can hardly make slits narrow enough to exhibit interference maxima of detectable width. Although X-ray interference is widely used, the essential diffracting apparatus is rarely man-made. Instead we employ the regular atomic patterns of crystals. For gamma rays, not even crystals will work as diffraction obstacles; only the nucleus of a single atom can be employed, and the procedure has become very indirect.

We have no choice; electromagnetic radiation of every kind must be understood from two points of view. If we wish to view the overall pattern it makes in space or time, we can do so by the theory of electromagnetic waves described in Chapter 29. We can design antennas, lenses, and gratings. They all function just as the wave theory predicts. But we must admit that we cannot say exactly when or where the energy of a photon will be transferred to the electron. The wave theory gives us only the probability of finding the photon. It tells only the probability that a photon will transfer its energy to a given object within a definite time. Here we are subject to the laws of chance. The photon cannot be pinned down by any wave prediction. In an interference experiment a particular photon may occur in any bright bar, and we cannot trace its path through the slits of the apparatus. But wherever and whenever it interacts, there and then we can be certain that energy is precisely conserved.

In the older mechanics and classical physics, we thought that we could know the position and velocity of a particle and at the same time tell when and where it was going to hit. That was naïve, and the study of the photon was the hard school in which physicists became more sophisticated in the ways of the world. For photons at least, we cannot do any more than assign probabilities to the time and place of arrival. These probabilities follow from the associated wave properties of light. We can therefore use only part of the ideas that we formerly associated with particles and their motions, and at the same time we must add new ideas from the wave picture and the probability interpretation to fill out our description of what actually happens.

Sometimes only one of the older pictures is needed. When many photons take part in any detectable change, it is the wave picture alone. On the other extreme, when a single photon is so energetic that it can easily be observed, and when it has as well so short a wavelength that interference studies are nearly impossible, then the photon picture is sufficient. But the real nature of light is more subtle. It shows us both wave and photon behavior. We have therefore been forced to make a new picture of light, one which combines some aspects of wave behavior as we previously understood it with some aspects of photon or particle behavior. This new picture is forced upon us, for this is the way the world is. Newton's corpuscular picture of light is by no means restored by the photon. It is no more true than the pure electromagnetic picture, with its waves of continuously distributed energy. Light is to be understood fully only by a novel scheme of the kind we have tried to outline.

QUESTIONS

1.* How would the two graphs in Fig. 31 – 2 look if the electrons were ejected according to the wave model of light?

2. In a certain camera, at dusk, the light that gets through the lens is so dim that in 1/100 second one photon on the average lands in 10 mm² of the film. If you want to take a finely detailed, sharp picture, what should you do?

3. Suppose, in Fig. 31 – 3, that there are *N* bullets/sec in the stream entering the region containing the set of targets, and that there are 0.80 *N* bullets/sec in the stream after it has covered a distance of *X* cm in the region.

   (a) How many bullets/sec will there be in the stream after 2*X* cm? After 3*X* cm?

   (b) Is this behavior consistent with that of light? (See Question 20, Chapter 5).

4. With a motion-picture camera you could determine the time at which the charge on each speck was changed for a large number of specks in the apparatus of Fig. 31 – 1. What relationship would you expect between the number of specks on which the charge changed and the time after the light was turned on? Does the data given in Fig. 31 – 2 support your prediction?

5.* A surface ejects electrons when hit by green light, but none when hit by yellow light. Do you expect that electrons will be ejected when the surface is hit (a) by red light? (b) by blue light?

6. Some photographic materials can be handled safely in red light but are spoiled instantly when white light is turned on. How do you account for this?

7. Suppose that a photograph similar to Fig. 31 – 4 had been made with X rays of wavelength about 0.1 angstrom. In what respects would you expect the new photograph to resemble Fig. 31 – 4? In what respects would it be different?

8. Assume that the entire negative plate in the cell illustrated in Fig. 31 – 5 is always illuminated by a constant source of light capable of producing a photoelectric current. What would be the effect on this current of

   (a) doubling the area of the plates?

   (b) halving the distance between the plates?

9.* In Fig. 31 – 7, which points on the curve were obtained by using violet light?

10.* What is the minimum energy, in electron volts, needed for photons to eject electrons from the lithium surface used in Fig. 31 – 7?

11. A good mirror reflects about 80 percent of the light which hits it. How could you determine whether 20 percent of the photons were not reflected, or whether all of the photons were reflected but each with 20 percent less energy?

12. Red light has a wavelength of about 6500 angstroms.

    (a) Estimate its frequency.

    (b) Estimate the energy in joules of one photon of red light.

    (c) Estimate the energy in electron volts of one such photon.

    (d) Find out approximately the longest and shortest wavelengths you can see, and compute the energy of the corresponding photons.

13. Light of 5000-angstrom wavelength illuminates a surface. What voltage is needed to stop all the photoelectrons emitted from the surface if their binding energy to the surface is 2.0 electron volts?

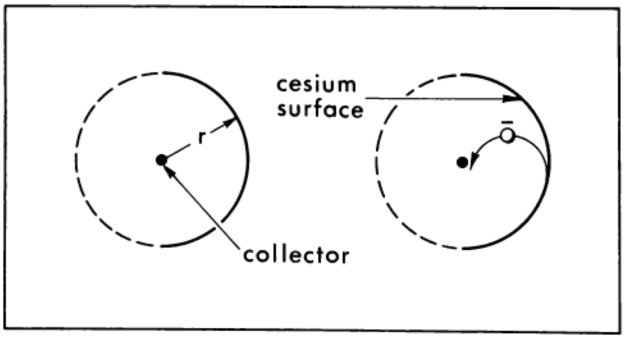
14. Sketch the graph of Fig. 31 – 6, then upon it sketch graphs that will show the effect of
   (a) increasing the intensity of the light source.
   (b) decreasing the wavelength of the light.
   (c) increasing the potential difference between the electrodes.

15. In Fig. 31 – 6, the current reaches a maximum value as the receiving plate is made positive. How could this maximum current value be changed?

16. (a) What is the ratio of the energy of a photon of wavelength 4500 Å (blue light) to the energy of a photon of wavelength 6300 Å (red light)? (b) What is the ratio of the maximum energy of electrons emitted from a surface with a binding energy of 1.8 electron volts using the same red and blue lights?

17. A cesium photocell (threshold energy of 1.96 ev) has a photoelectric surface in the shape of a half cylinder of radius 1.0 cm. (See Fig. 31 – 10.) The collector is a wire along the axis of the cylinder. What would be the strength of a magnetic field, parallel to the axis of the cylinder, which would prevent any electrons emitted from the surface from reaching the collector if green light of wavelength 5000 Å were used to illuminate the surface?

18. (a) Take a pack of ordinary playing cards (from which the jokers are removed), thoroughly shuffle them, and draw a card. After noting which card it is, return it to the pack, shuffle, and draw again. After ten draws, note how many times each of the following were drawn:
   (I) A face card.
   (II) A card on which a 3 is printed.
   (III) A card on which an even number is printed.
   (b) Make a bar graph, after the fashion of Fig. 31 – 8, to display the results of ten draws.
   (c) Make a bar graph to display the results you would expect to obtain as the result of about 5200 draws.
   (d) In what respects are your graphs like those of Fig. 31 – 8?



31–10 For Problem 17.

19. In the three-month experiment described in Section 31 – 5, G. I. Taylor estimated the energy of the light reaching his photographic plate. He found it was 5×10^{-13} joule per second. (He obtained that value by comparing the average blackening of the plate with that produced in 10 seconds by a candle 2 yards away, without any absorbing screens of smoked glass.) From that estimate, calculate the average distance between photons, as follows:
   (a) Assume the wavelength of the useful light was about 5000 angstrom units. What energy did each photon carry?
   (b) Calculate from the given flow of energy the average time that elapsed between the arrival of one photon and the next.
   (c) From the average time, calculate the average distance between one photon and the next in the beam of light.
   (d) If you were asked, "How many photons are in the box at any chosen instant?" you would have to answer "None" most of the time, but on looking again and again and again, you should expect to "see" one photon in the box about once in so many trials. How many? Assume the box was 1.2 meters long.

20. Use the language of photons to describe how the pattern of Fig. 9-5 is formed with very dim light.

21. An ordinary flashlight has a bulb which draws 0.25 amp at 3.0 volts, and concentrates the emitted light into a beam about 10 cm^2 in area. Assume that 1.0% of the energy consumed by the bulb is turned into visible light of wavelength 5.8×10^3 angstrom units.
   (a) How many photons are there in each meter along the axis of the beam?
   (b) How many photons per second arrive on each square centimeter of a surface illuminated by this flashlight?

22. A typical FM radio station broadcasts with a wavelength of 3 meters. Estimate:
   (a) the energy in one photon of this radiation.
   (b) the number of photons broadcast per second if the radiated power is 10 kilowatts.
   (c) roughly the number of photons received by a receiving set during one vibration of the letter "s" in the audio pattern, if the receiver is

100 km from the station. Make the following simplifying assumptions:

(i) the radiation spreads uniformly in all directions along radii from the source so that an inverse-square law holds for energy flow (This is not true for any ordinary antenna.)

(ii) the receiving antenna collects radiation from an area of 1 meter^2 perpendicular to the line of travel of waves.

(iii) the sound of the letter "s" involves air oscillations of frequency about 4000 cycles per second.

(d) Will you ordinarily notice the individual photons in receiving the letter "s"?

23. In a particular X-ray tube electrons hit a target after they have been accelerated by a potential difference of 20,000 volts. As they decelerate to a stop in the target, a few of them emit X-ray photons.

(a) Why is there a definite minimum wavelength of photons observed from this tube? (This limit is known as the Duane-Hunt limit.)

(b) Calculate that minimum wavelength.

32

Speed, Energy, and Mass

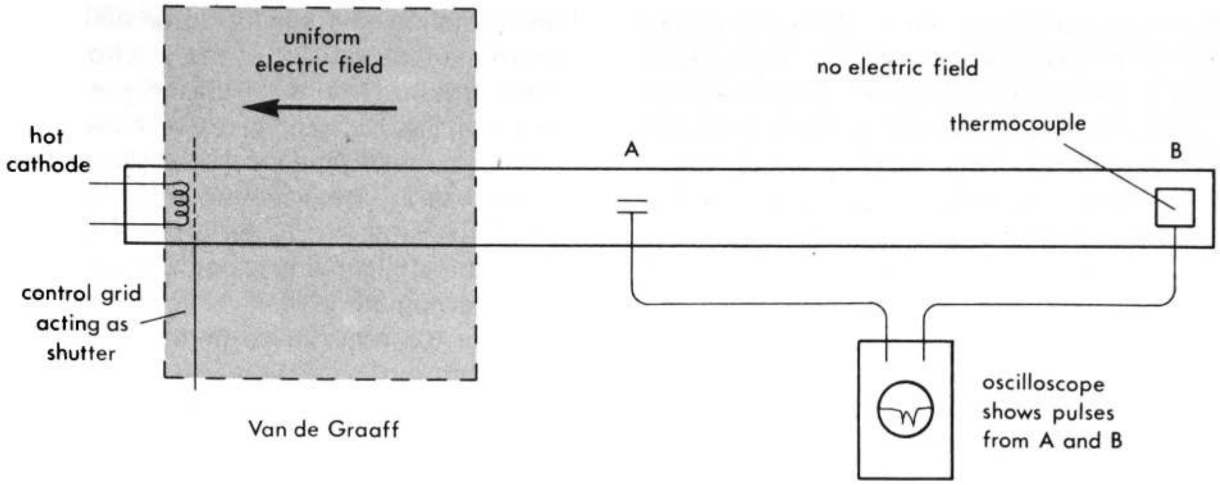
NEWTON'S LAW of motion says that a body will accelerate uniformly under the action of a constant force. Thus, the speed of a body initially at rest should increase linearly with time: $v = (F/m)t$. If in this time the body covers a distance d, then its kinetic energy will be $E_K = Fd = \frac{1}{2}mv^2$. These relations have been amply confirmed in the domain of low speeds, that is, for speeds small compared with the speed of light. They certainly do not hold for photons: photons move with speed c, independent of their energy.

On the basis of the ideas developed in Chapter 30, we have reason to suspect that as we work with speeds approaching the speed of light, the Newtonian relations given in the previous paragraph have to be modified. The factor $1/\sqrt{1 - v^2/c^2}$ appears in all our transformation equations, and this factor can hardly have a physical meaning if $v > c$. All the kinematics we have developed in Chapter 30, then, breaks down for speeds greater than c. Let us, therefore, investigate the question experimentally and see if the results can guide us into a modified relativistic dynamics that does not predict impossibly high speeds and accounts for the behavior of photons.

32–1 THE ULTIMATE SPEED

We can hardly use ordinary macroscopic objects in our investigation, because exceedingly high energies would be required to provide speeds close to that of light. Our problem is greatly simplified if we use the lightest available bodies—electrons. For example, electrons accelerated through a potential difference of 2500 volts in a cathode-ray tube are moving at a speed of 0.1 c.

The effect is much more dramatically illustrated if we can use accelerating potentials of about a million volts. A Van de Graaff accelerator at the Massachusetts Institute of Technology was used as a source of high voltage for an experiment of this type. The general arrangement of the

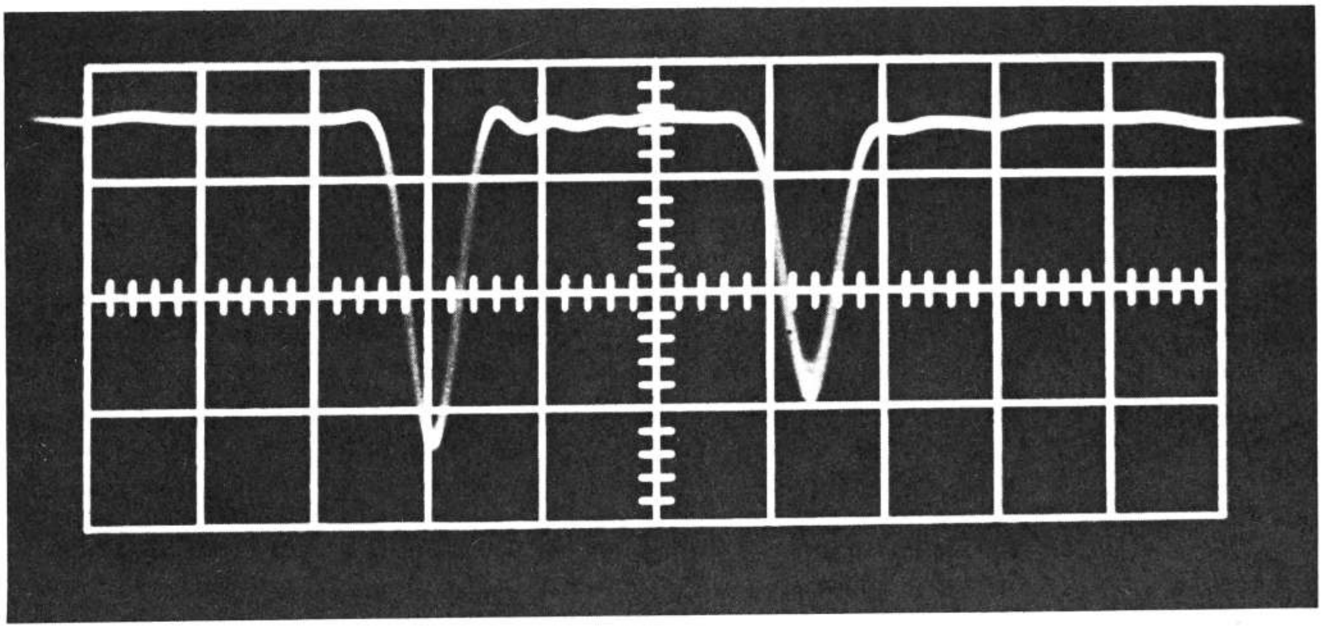


32–1 The general arrangement of the experiment performed at M.I.T. by William Bertozzi. The electrons are accelerated in a uniform field on the left and timed between A and B by the oscilloscope.

apparatus is shown in Fig. 32 – 1.

Electrons from a hot cathode are accelerated by a large difference of potential maintained by the Van de Graaff. They emerge from the Van de Graaff into a long tube which has no electric field in it. The electrons move down this tube without accelerating, and their speed is found by timing their flight from *A* to *B*. For accurate measurement it is essential that the electrons be emitted in bursts whose duration is short compared with the time of flight. A shutter that will open for extremely short time intervals has been made by placing a grid close to the cathode and keeping it charged negatively until we wish to emit a burst of electrons. We then apply a sharp pulse of positive charge to the grid, and a short burst of electrons will be released.

The time of flight is indicated by an oscilloscope; when the bursts of electrons pass *A* and *B*, electrical pulses are generated and these appear as spikes on the oscilloscope face (Fig. 32 – 2).



32–2 The pulses generated at A and B as seen on the oscilloscope. The sweep speed of the oscilloscope was set in such a way that the distance between the small divisions on the horizontal axis corresponds to 2×10^{-9} sec. The pulses shown here are from the run with an accelerating potential of half a million volts.

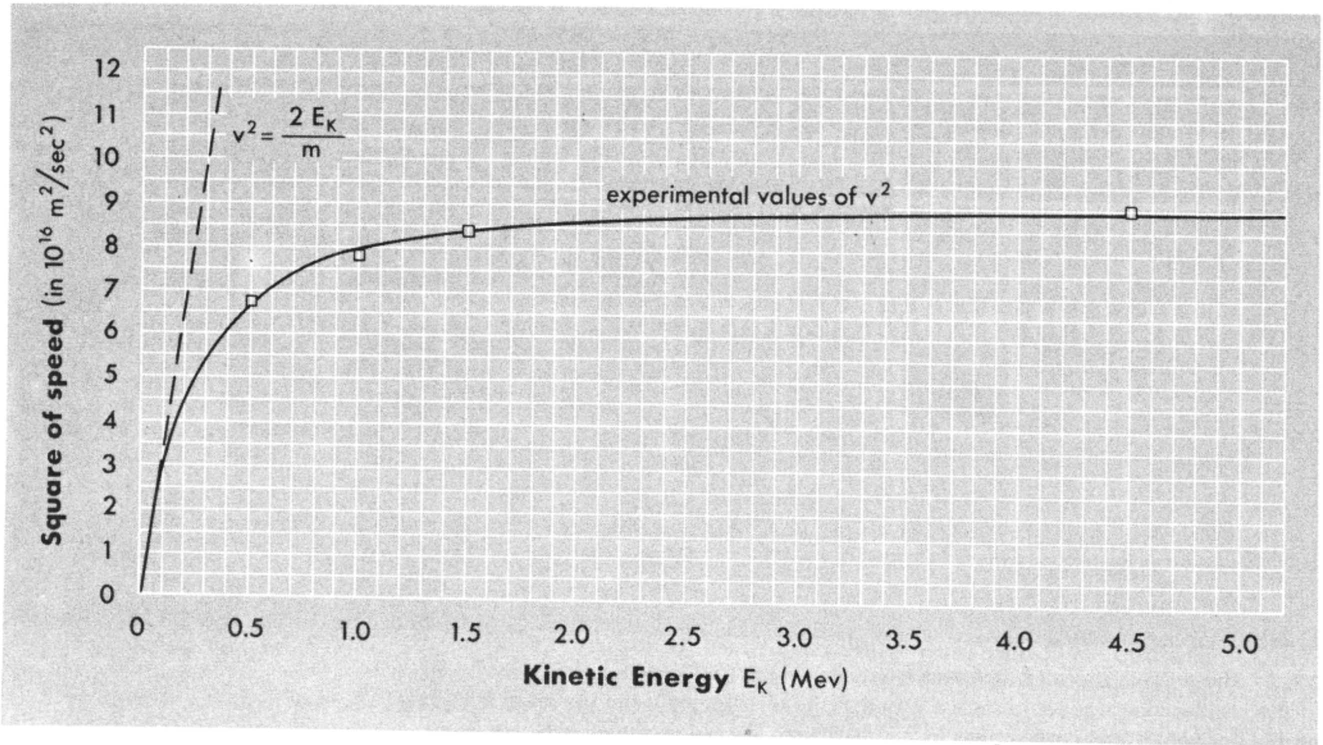
All the precautions taken in determining the speed of the electrons assure its correctness. But can we be equally sure of the correctness of the values of kinetic energy of the electrons? The kinetic energy of a particle which starts from rest equals the work done on it, that is, the force times distance. The distance is easily measurable. If the speed of the electron would remain low, we could calculate the force by dividing the potential difference between the two terminals of the Van de Graaff by the distance between them: $F = qV/d$ and $E_K = (qV/d)\,d = qV$. Does this relation hold also when the electron moves at speeds close to the speed of light?

To check up on this point, an aluminum disc containing a thermocouple was placed near *B* (Fig. 32 – 1). The kinetic energy of the impinging electrons was converted into thermal energy of the disc, and the rise in temperature was measured. From the rise in temperature and the number of electrons hitting the cylinder, the kinetic energy of each electron was calculated and found to agree with the potential energy. Since the energy transferred to each electron is $E_K = Fd$, and we are now sure of knowing both E_K and *d*, we have also shown that in our experiment, where the force is at all times in the direction of motion, the electric force acting on the electron is independent of the electron's speed.

The striking results of this experiment are shown in Fig. 32 – 3. Since the Newtonian relation $E_K = \frac{1}{2}mv^2$ predicts that v^2 should be proportional to E_K, we have plotted v^2 (in units of 10^{16} m²/sec²) against E_K (in units of 10^6 ev = 1 Mev). The straight line going up without limit is the Newtonian relation. Note that even at an energy of 0.5 Mev, the experimental point deviates from this prediction. As we go to higher energies, v^2 clearly levels off at 9×10^{16} *m²/sec²* $= c^2$, and the difference from the classical prediction becomes enormous. This result is quite general: never has a particle been found to have a speed larger than *c*.

32–2 SPEED AND KINETIC ENERGY

Experiments involving speeds close to the speed of light do not disprove the laws of classical physics; they only point out that these laws cannot be simply extrapolated without bounds, but must be modified at high speeds because of the special role played by the speed of light. When we look for the extensions of the physical laws



32–3 A plot of v^2 vs. E_K for electrons. The dashed straight line represents the classical prediction; the curve goes through the experimental points.

into the relativistic domain, we must not lose sight of classical physics, because all the modifications which we introduce must become negligibly small at low speeds. The principle applied here is fundamental in the development of any theory. (Compare Bohr's Correspondence Principle in Section 33 – 5 and the Relativistic Addition of Velocities, Chapter 30 Section 10.) We shall use it in finding the general relation between speed and kinetic energy.

We begin by redrawing Fig. 32 – 3 in the following way: we plot the kinetic energy on the vertical axis and on the horizontal axis we plot v/c (Fig. 32 – 4). The interchange of the axes has the advantage of presenting the relation between kinetic energy and speed in the form with which we are most familiar: $E_K = \frac{1}{2}mv^2$, which can be written as $E_K = mc^2(\frac{1}{2}v^2/c^2)$. We plot v/c on the horizontal axis because we found it useful to express deviations from classical behavior in terms of v/c.

Notice the three experimental points at 0.5, 1.0, 1.5 Mev. The point at 4.5 shown in Fig. 32 – 3 is not shown here, but the trend is clear: the curve goes straight up as v/c approaches 1. The low-energy end of the curve is the classical region; as in Fig. 32 – 3, the experimental graph overlaps in this region with the graph of the parabola $E_K = mc^2(\frac{1}{2}v^2/c^2)$.

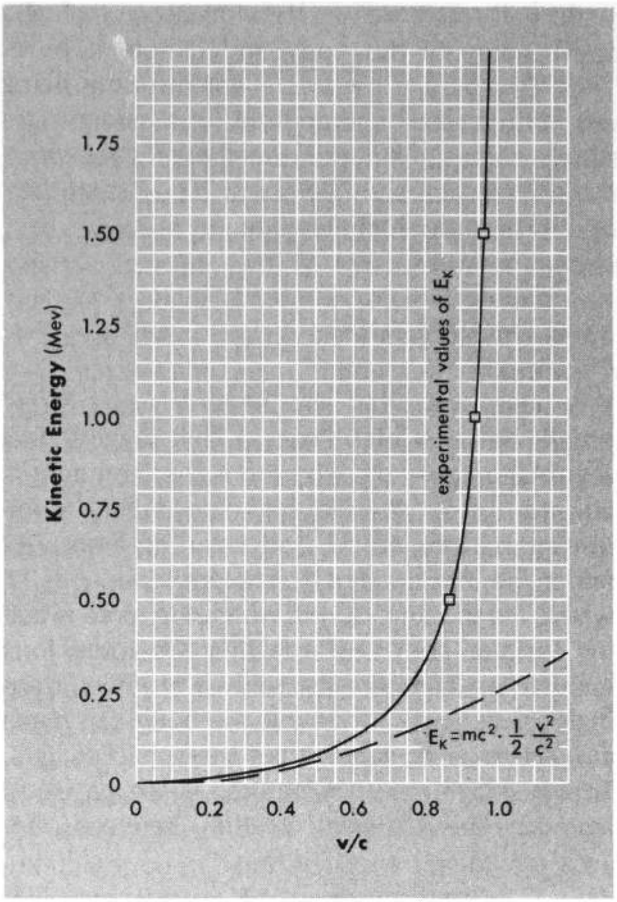
Our question is now the following: which function of v/c behaves like $\frac{1}{2}v^2/c^2$ when $v/c \ll 1$ and goes rapidly to infinity as v/c approaches 1? We have already met one which has these characteristics. In the discussion of the Michelson-Morley experiment (Chapter 30, Section 4) we encountered the function $\left(\frac{1}{\sqrt{1-v^2/c^2}} - 1\right)$

We have shown in the Box of that section that for $v/c \ll 1$,

$$\left(\frac{1}{\sqrt{1-v^2/c^2}} - 1\right) \approx \tfrac{1}{2}v^2/c^2$$

We can easily see that as v/c approaches 1, the function goes up extremely steeply. Of course, there are many functions which fulfill these general conditions, but a comparison of the general appearance of Fig. 32 – 3 and Fig. 30 – 18 strongly suggests that the dependence of kinetic energy on speed is given by

$$E_K = mc^2\left(\frac{1}{\sqrt{1-v^2/c^2}} - 1\right)$$



32–4 A plot of E_K vs. v/c for electrons. The dashed curve represents the classical prediction.

To test the correctness of this expression we shall use it to calculate E_K for some value of v/c and compare this calculated E_K with the value obtained from the experimental curve. For example: $mc^2 = 0.9 \times 10^{-30} \times 9 \times 10^{16} = 8 \times 10^{-14}$ joule. Recalling that 1.6×10^{-19} joule $= 1$ ev (see Section 25 – 7), we find

$$mc^2 = 0.5 \times 10^6 \text{ ev} = 0.5 \text{ Mev}$$

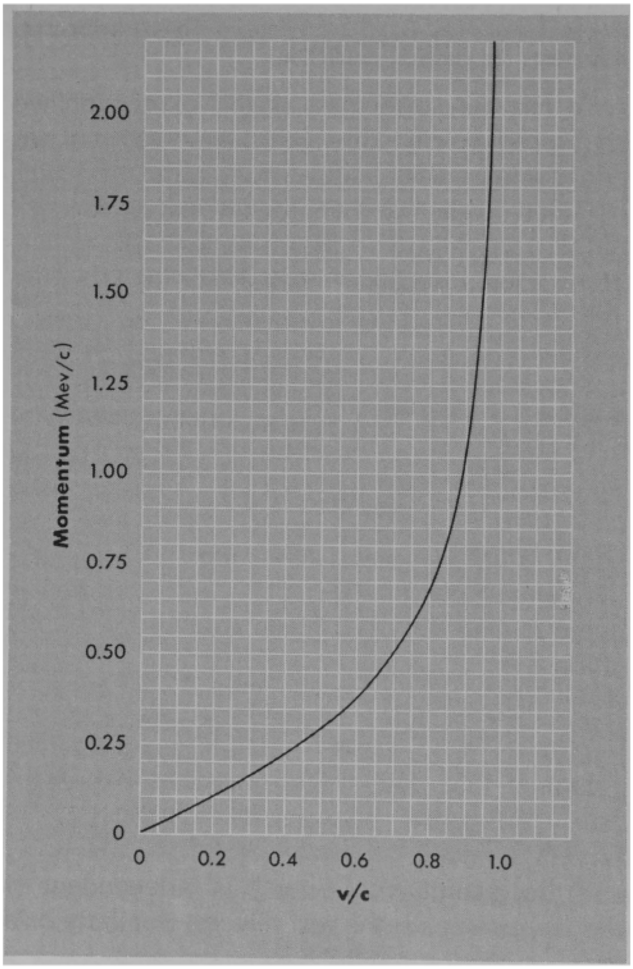
For $v/c = 0.87$ we find a kinetic energy of $E_K = 0.5\left(\frac{1}{\sqrt{1-0.87^2}} - 1\right) = 0.5$ Mev. This value is in complete agreement with the experimental curve. In fact, the experimental curve in Fig. 32 – 4 is given exactly by the relation

$$E_K = mc^2\left(\frac{1}{\sqrt{1-v^2/c^2}} - 1\right)$$

Notice that the kinetic energy is expressed as a difference between two terms: $E_K = (mc^2/\sqrt{1-v^2/c^2}) - mc^2$. It is convenient to call the quantity mc^2, which is independent of the body's speed, the rest energy. Similarly, we shall call $mc^2/\sqrt{1 - v^2/c^2} = E_K + mc^2$ the total energy of the body and denote it by E. Thus $E = mc^2/\sqrt{1 - v^2/c^2}$. For the time being the terms total energy and rest energy are simply names given to the mathematical expressions. We shall see in Section 32 – 7 that they also have a physical meaning.

32–3 MOMENTUM

We have seen that the relation $E_K = \frac{1}{2}mv^2$ breaks down at high speeds. This suggests that the relation $p = mv$, where p is the momentum, will also break down at high speeds for the following reason. For a body starting from rest under the influence of a constant force, $p = Ft$, where t is the time during which the force is acting on the body. By applying the force long enough we can give the body a momentum as high as we wish. If $p = mv$, the body can reach any speed, even larger than the speed of light. This is not the case in nature, as we have just learned. The classical relation between momentum and speed also requires modification at high speeds.



32–5 **A plot of p vs. v/c for electrons.**

We begin by looking first at the dependence of p on v when v is very close to c. Think, for example, of the electrons which went through a potential difference of 4.5 Mev. (See Fig. 32 – 3.) Except at the very beginning of their trip they moved with almost constant speed, the speed of light. Therefore, the electric force acted on them during a time $t \approx d/c$, giving them a momentum $p = Ft = Fd/c = E_K/c$. But in this case $E_K \approx E$, the total energy, because $E_K = 4.5$ Mev and the rest energy of the electron $mc^2 = 0.5$ Mev, which is only about 10 percent of E_K. So here $p \approx E/c = mc/\sqrt{1 - v^2/c^2}$. This expression shows that the momentum can become as large as we wish while the speed remains less than the speed of light. It is valid only at very high speeds. At low speeds we have the familiar expression $p = mv$. The general expression for the momentum as a function of speed which we wish to find must, therefore, reduce to $p = mv$ when $v \ll c$ and to $p = mc/\sqrt{1 - v^2/c^2}$ when v is very close to c. We see quite readily what the general expression must be

$$p = mv/\sqrt{1 - v^2/c^2}$$

Indeed, for $v \ll c$, $1/\sqrt{1 - v^2/c^2} \approx 1$, and $p = mv$ as required, and when $v \approx c$, we can replace v in the numerator by c because only the denominator is sensitive to minute changes in v. Fig. 32 – 5 shows a graph of p vs. v/c for electrons.

So far we have discussed only the magnitude of the momentum; but momentum is a vector quantity. At low speeds $\vec{\mathbf{p}} = m\vec{\mathbf{v}}$. The factor $1/\sqrt{1 - v^2/c^2}$ is independent of direction. Therefore, the relativistic expression for momentum must be $\vec{\mathbf{p}} = m\vec{\mathbf{v}}/\sqrt{1 - v^2/c^2}$.*

The expression $p = mv/\sqrt{1 - v^2/c^2}$ has the same factor $1/\sqrt{1 - v^2/c^2}$ as the expression for what we called the total energy in the preceding section:

$$E = mc^2/\sqrt{1 - v^2/c^2}$$

This suggests that E and p can be related in a

---

*In many textbooks the mass of a particle is denoted by m_0 and called the "rest mass." Another mass, the "relativistic mass" is denoted by m and defined by the relation $m = m_0\sqrt{1 - v^2/c^2}$. In this notation the expression for the momentum looks like the classical one: $\vec{\mathbf{p}} = m\vec{\mathbf{v}}$, but m, nevertheless, is now a function of the velocity. The total energy is then written as $E = m_0c^2/\sqrt{1 - v^2/c^2} = mc^2$. But the kinetic energy *cannot* be expressed as $\frac{1}{2}mv^2$. Thus, relativistic dynamics is not reduced to classical dynamics simply by replacing m_0 by $m = m_0/\sqrt{1 - v^2/c^2}$.

simple way. Indeed, by multiplying and dividing p by c^2 we find

$$p = mc^2v/c^2\sqrt{1 - v^2/c^2} = Ev/c^2$$

When energies are expressed in units of Mev, it is convenient to express momenta in units of Mev/c (read "Mev over c"). A particle of total (kinetic + rest) energy E expressed in Mev, has a momentum of $(E/c)\cdot(v/c)$ expressed in Mev/c. For example, a particle with total energy 1.7 Mev and speed $v = 0.95c$ has a momentum $p = 1.7 \times 0.95 = 1.6$ Mev/c.

Let us compare the two expressions $p = mv/\sqrt{1 - v^2/c^2}$ and $p = Ev/c^2$. The first can be applied only to situations where the mass is known and the speed is less than the speed of light. The latter condition is there because $1/\sqrt{1 - v^2/c^2}$ becomes meaningless for $v = c$. Using the second expression does not require the knowledge of the mass. It may even apply to particles which have no mass as long as they have energy and speed. Furthermore, since this expression does not contain $1/\sqrt{1 - v^2/c^2}$, but only v, it can also be applied when $v = c$. This suggests that we try to use it for photons.

32–4 THE MOMENTUM OF PHOTONS: THE PRESSURE OF LIGHT

The particle model of light discussed in Chapter 5 suggested that light exerts a pressure when it is reflected or absorbed. True, we encountered serious difficulties with the early particle model and had to discard it. But, in the preceding chapter, many of the basic aspects of the particle model became useful again.

The new light particle, the photon, differs from the old light particle in a fundamental way. In Newtonian mechanics the most basic property of a particle is its mass. When it moves, its momentum and kinetic energy are proportional to its mass. If there is no mass, there is no momentum and no kinetic energy. But from Section 32 - 1 we know that a particle with mass cannot move with the speed c. The photon, which does move with the speed c, must therefore have zero mass. We know from experiment that a photon has an energy $E = h\nu$ where h is Planck's constant and ν is the frequency of the light. (See Section 31 - 3.) This energy is all kinetic energy, since a particle without mass has no rest energy. The relation $p = Ev/c^2$ derived in the preceding section enables us now to find the momentum of a photon of energy E:

$$p = Ec/c^2 = E/c$$

This simple proportionality between momentum and energy means that it is irrelevant whether a beam of light is made up of relatively few energetic photons or many photons of little energy. In other words, two beams of light of the same energy have the same momentum independently of the colors present in the beam.

When a light beam hitting a shiny surface at right angles is reflected, its momentum is reversed. The change of momentum of the reflector due to the reflection of a large number of photons of total momentum p is $\Delta p = 2p$. If these photons hit the reflector in a time Δt, the change of momentum per unit time is

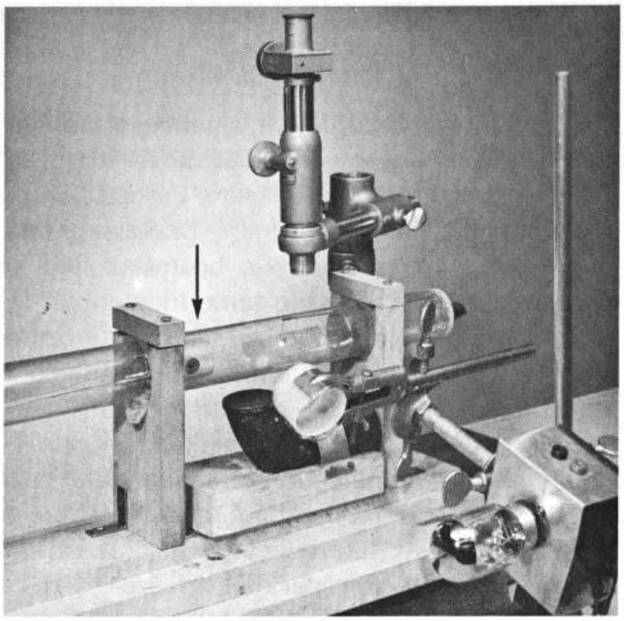
$$\Delta p/\Delta t = 2p/\Delta t = 2E/c\Delta t$$

and this is the force exerted by the light beam on the reflector.

E. F. Nichols and G. F. Hull checked this relation experimentally. They measured the force exerted by an intense light beam on a mirror mounted on a very sensitive torsion balance. Then they found the energy carried by the light beam per unit time by shining the light on a silver disc coated with lampblack. The lampblack absorbed almost all the light which fell on it, and Nichols and Hull measured the resulting rise in temperature of the silver disc. Taking into account the small amount of light that was reflected, the rise in temperature gave them a good determination of the energy of the light beam. Finally, they compared the force measured directly with the force calculated from the energy of the beam. They made their experiment three times: using white light, white light with a ruby glass filter, and white light with a water filter. The ruby glass removed most of the higher frequencies from the green to the ultraviolet; the water removed the infrared part of the spectrum. Their results are shown in Table 1.

TABLE 1

	FORCE IN NEWTONS ON 1 CM² OF MIRROR	
LIGHT BEAM FILTERED BY	MEASURED DIRECTLY	CALCULATED FROM ENERGY
No filter	7.01×10^{-10}	7.05×10^{-10}
Ruby glass	6.94×10^{-10}	6.86×10^{-10}
Water	6.52×10^{-10}	6.48×10^{-10}

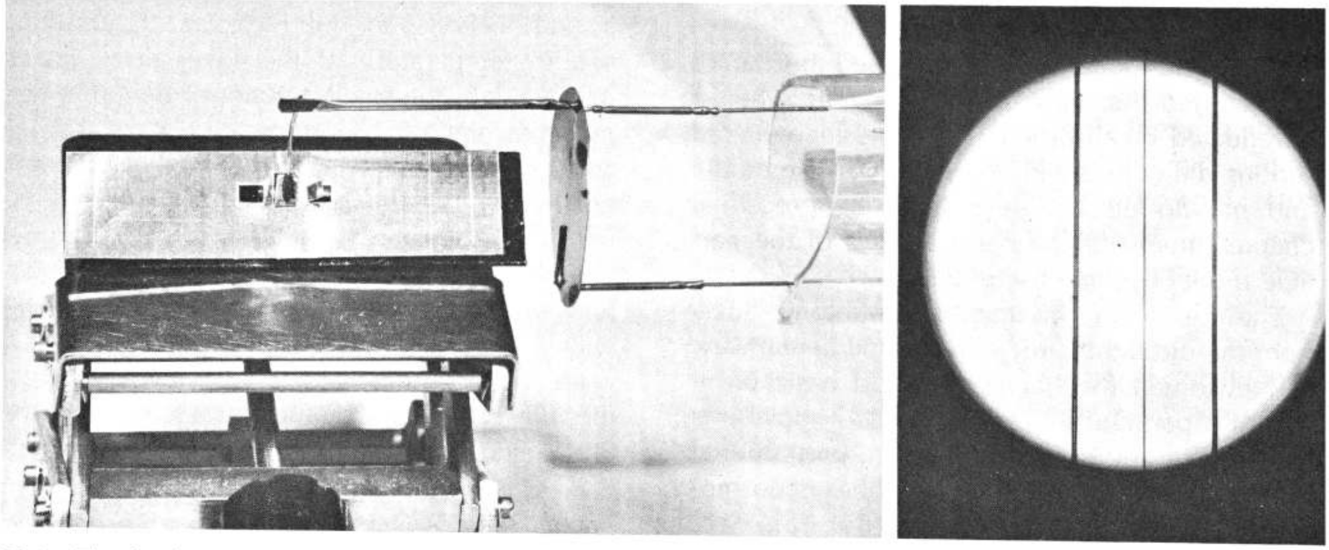


32–6 A general view of the apparatus used to measure the pressure of light. The torsion balance is mounted in an evacuated tube. The can for measuring the energy of the light beam is indicated by the arrow. Below the vacuum tube is a magnet for steadying the aluminum strip by inducing currents in it while it is in motion.

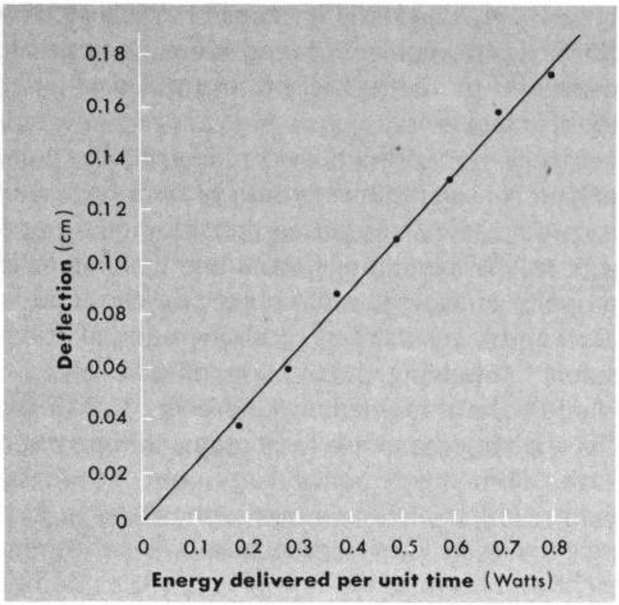
The close agreement between the measured force, which is the rate of momentum transfer, and the force calculated from the energy confirms the momentum-energy relation used. Since blue photons are more energetic than red ones, we see also that this relation does not depend on the energy of the photons.

A modern version of the same experiment performed by A. P. French is illustrated in Figs. 32 – 6 and 32 – 7. The torsion balance is made by hanging a strip of light aluminum foil on a fine glass fiber inside a glass tube pumped out to a high vacuum. A lens concentrates an intense spot of light near one end of the strip. The impact pushes the end of the strip back, twisting the fiber, and the deflection is measured with a microscope focused on a mark at the other end of the strip.

The energy carried by the light is measured by sending it through a small hole into the blackened interior of the can indicated by the arrow in Fig. 32 – 6 and there converting it to thermal energy. The can also contains an electric heater which can be adjusted until it produces the same rise in temperature as the light does. The results of the experiment are shown in Fig. 32 – 8. Like those of Nichols and Hull, they confirm the relation $p = E/c$ for photons.



32–7 The aluminum strip which forms the horizontal arms of the torsion balance is shown supported on a laboratory jack to keep it from fluttering in air currents while the picture is taken. During the experiment, the light beam is aimed at the cross ruling at the left end of the strip while the microscope is focused on the right end (Fig. 32–6). The wire leading down to the aluminum strip is used in setting up the apparatus; it is not the suspension fiber, which is too fine to be seen in the photograph. This fiber can be seen in the photograph at the right, taken through a microscope; the two dark lines are microscope scale rulings 1 millimeter apart.



32–8 The deflection of the balance is proportional to the force exerted on it. The experimental points are in good agreement with the straight line predicted by the theory using p = E/c.

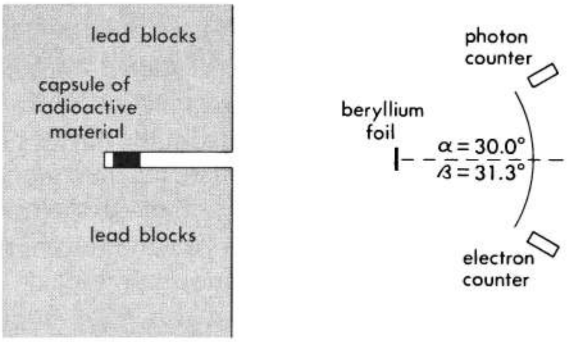
32–5 COMPTON SCATTERING

In the preceding section we used the pressure of light to study the momentum of photons and check up on the relativistic expressions for energy and momentum. In that experiment a light beam containing a large number of photons was reflected from a mirror made of a large number of atoms. The mirror as a whole barely moved and could, therefore, be adequately described by classical mechanics. To give further support to the relativistic expressions for energy and momentum, we wish to study collisions between one photon and one electron. If the photon is energetic enough, the electron may be pushed sufficiently hard to recoil with a speed comparable to the speed of light. Thus both photon and electron will have to be described relativistically.

Energetic photons can be obtained from an X-ray machine or a radioactive material. Electrons are, of course, present in all matter, where they are bound to the atomic nuclei. But as far as collisions with energetic photons are concerned, the outer electrons can be considered free and at rest.

Collisions between photons and electrons were first studied by A. H. Compton in 1923. He observed that the scattered photons had longer wavelengths and hence lower energies than the incoming photons. The observed relation between the increase in wavelength and the angle of scattering is just the one you would obtain by assuming that momentum and energy are conserved in the collision.

The recoiling electron was not observed in Compton's original experiment. In spite of the agreement of the observed and the predicted relation between the energy and the angle of scattering, it was felt by many that further experiments were needed to justify the application of conservation of energy and momentum to a situation so far removed from the domain of Newtonian mechanics, where these principles were first discovered. In particular, both the scattered photon and the recoil electron should be observed simultaneously. This was done by W. G. Cross and N. F. Ramsey with the apparatus sketched in Fig. 32 – 9. The geometry of the arrangement permits a very simple analysis which warrants a detailed discussion. Here the angles made by the scattered photon and the recoil electron with the direction of the incoming photon are equal within 4 percent. For convenience we shall approximate both by 30°.



32–9 A schematic diagram of the experimental setup for observing Compton scattering. A narrow beam of gamma rays was produced by inserting a capsule of radioactive material in a small, deep hole in a pile of lead blocks. A very thin foil of beryllium metal scattered the photons. Those which came off at an angle of 30.0° fell upon a crystal of anthracene, which glowed briefly as each photon entered it. The flashes of light were picked up by a photomultiplier and converted into short electrical pulses. Another anthracene crystal set at 31.3° on the opposite side of the beam of gamma rays detected the recoil electrons. An electronic coincidence counter received the two electrical pulses and recorded a count only if they arrived within an interval of 1.5 × 10^{-8} sec. Each count then represented a single collision between a photon and an electron.

If momentum is conserved in the collision, the vector sum of the momenta of the photon ($\vec{\mathbf{p}}_f$) and the electron ($\vec{\mathbf{p}}_e$) after the collision must equal the momentum of the incoming photon ($\vec{\mathbf{p}}_i$):

$$\vec{\mathbf{p}}_f + \vec{\mathbf{p}}_e = \vec{\mathbf{p}}_i$$

Writing this out for the components of the momenta perpendicular and parallel to $\vec{\mathbf{p}}_i$, we have

$$p_f \sin 30° - p_e \sin 30° = 0$$

and

$$p_f \cos 30° + p_e \cos 30° = p_i$$

From the first equation we find $p_f = p_e$. Recalling that $\cos 30° = \sqrt{3}/2$, we find from the second equation

$$p_f = p_e = \frac{1}{\sqrt{3}}\, p_i$$

Now we shall use the relativistic relation between momentum and energy to compare the energy before and after the collision. The energy of the incident photon was $E_i = 2.6$ Mev, hence $p_i = E_i/c = 2.6$ Mev/c, and $p_e = p_f = 1.5$ Mev/c. The energy of a photon of momentum 1.5 Mev/c is 1.5 Mev. To find the energy of the electron exactly we have to know its speed, since $p = Ev/c^2$. But a glance at Fig. 32–5 will convince us that an electron of this momentum moves with practically the speed of light. We shall, therefore, calculate the electron's energy as $E = pc^2/v \approx pc$. Hence the recoil electron, like the scattered photon, has an energy of 1.5 Mev. However, there is one important difference: the photon's energy of 1.5 Mev is all kinetic energy, whereas the electron's is kinetic energy plus rest energy.

We have found earlier that the rest energy of an electron is about 0.5 Mev. Therefore, the kinetic energy of the electron after the collision and the energy of the scattered photon add up to $(1.5 - 0.5) + 1.5 = 2.5$ Mev. This is very close to the energy of the incident photon of 2.6 Mev. Within the approximations which we made, the kinetic energy after the collision equals the kinetic energy before the collision. (An exact calculation using the correct angles and the correct speed reduces the difference considerably.) The observed phenomenon can be consistently described in terms of an elastic collision where momentum and kinetic energy are expressed by the relativistic relations developed in this chapter.

It may be worthwhile to reflect at this point on the considerable extension of our concept of a particle since we started the study of dynamics. At the beginning, the most characteristic property of a particle (an object we can hold in our hands) was its mass, both inertial and gravitational. Studying the motion of particles, we learned about momentum and energy. Now we are applying dynamics to photons, things which have momentum and energy, but no mass; things which always move with speed c. We treat them as particles because we can count them and because the conservation laws of momentum and energy apply also to their interactions with other, more conventional, particles.

Notice also that in the discussion of Compton scattering we have not mentioned the force of interaction between the electron and the photon. In classical dynamics the forces were very much in the foreground. We studied the relation between force and rate of change of momentum; we also studied how forces between bodies change as a function of their separation. How does the force between the photon and the electron vary as a function of separation? There is no way of following the electron and photon through the collision process; we can only determine the initial and final situations. This is why only momentum and energy are at the center of our discussion of relativistic particles.

32–6 THE ELECTRON-POSITRON ANNIHILATION

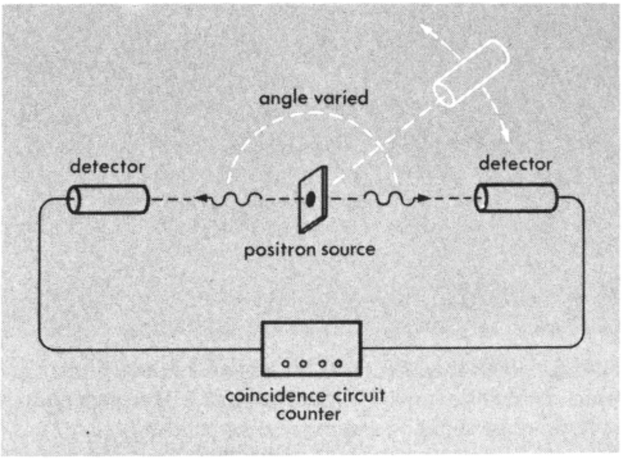
When light is emitted, photons are created; when light is absorbed, photons are destroyed or annihilated. When light is emitted by atoms, the energy of the photon is supplied by the atom; after the creation of the photon the atom has less internal energy than before, but in other respects it is the same kind of atom. We shall now discuss a case where photons are created by a more drastic process, a process in which the emitting particles themselves disappear.

In this experiment electrons are attracted not to positively charged nuclei, but to positively charged electrons, or positrons for short. Positrons were first noticed in 1932 by C. D. Anderson while he was measuring the energy of electrons which come down from the upper atmosphere as part of the cosmic rays. He had a cloud chamber placed in a magnetic field so that the electrons moved in circular trajectories across it, and from the radii of the circles he calculated the momenta of the electrons. Occasionally he saw a track which curved in the opposite way, as though it had been traced by a positively charged particle. Except for the direction of their curvature, these odd tracks looked exactly like the tracks of electrons: they stretched as far across the cloud chamber; the number of water drops along them was the same. There was every indication that these tracks had been made by particles of the same mass as electrons, but carrying a single, positive elementary charge. Shortly after Anderson's discovery, positrons were also observed to be emitted by some radioactive elements.

What happens to positrons once they are created? Since positrons are just like electrons except for the sign of their charge, we expect them to be slowed down in matter and finally come to rest. This has been demonstrated many times in cloud-chamber photographs and by other means. In material of high atomic number such as gold, positrons, like electrons, have a very short range. Placing a positron-emitting source on a small gold plate will guarantee that the positrons emitted into the plate will come to rest very close to where they were created. A positron at rest, or almost at rest, will attract an electron with the same force as a proton does, because they have the same electric charge. It seems then reasonable to expect that a positron-electron pair will form something like a "hydrogen atom." This indeed happens occasionally, but it is never the end of the story.

A Geiger counter placed not far from the source and the small gold plate reveals the presence of some energetic particles coming from the gold plate. Careful measurements have shown that these particles are photons, whose energy is always very close to 0.5 Mev. This amount of energy is equal to the rest energy mc^2 of either a positron or an electron. Is it possible that the positron by itself simply decayed into a single photon? A positron at rest has no momentum; a photon of 0.5 Mev has a momentum of 0.5 Mev/c in the direction of its motion. Thus, momentum would not be conserved in this process. Furthermore, the positron carries one elementary electric charge, and the photon carries none; so electric charge, which so far seems always to be conserved, will not be conserved here. It will be much easier to account for the creation of photons of 0.5 Mev if we assume that both the positron and an electron disappear together, giving off two photons. If the two photons go off in opposite directions, their total momentum is zero, just like the momentum of the positron-electron pair. The total electric charge of the pair is zero, so that electric charge also is conserved.

Back in 1932, when positrons were first observed, there was reason to believe on theoretical grounds that a positron finds its end by meeting an electron and disappearing, creating two photons. The process has been confirmed by the following experiment. Two detectors were placed near the source of the positrons and the plate in which they were stopped (Fig. 32 – 10). The detectors were connected through a coincidence circuit which registered a count only when both detectors recorded photons simultaneously. The counting rate did not depend on the angle formed by the source and the two counters, as long as that angle did not approach 180°. Near 180° the counting rate increased considerably, thus showing that when a photon is emitted in one direction, another photon is emitted in the opposite direction.



32–10 The arrangement used to demonstrate the simultaneous creation of two photons.

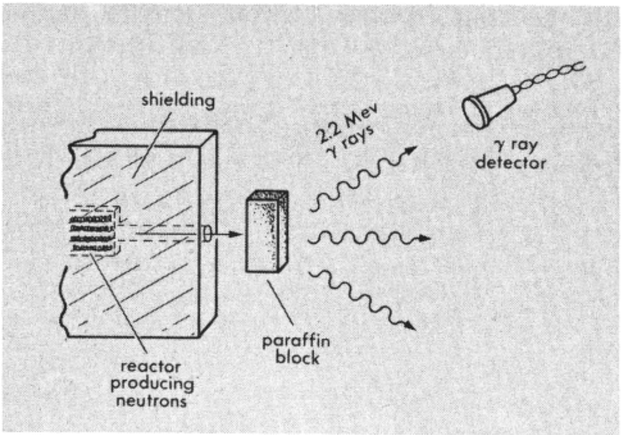
On the basis of this and other experiments, we now believe that an electron and a positron almost at rest annihilate each other while creating two photons. Energetic photons are often referred to as gamma (γ) rays. Hence we write

$$e^+ + e^- \rightarrow 2\gamma$$

Electric charge and momentum are conserved in the process, but what about the mass and kinetic energy? The electron-positron pair has a mass $2m$ and no kinetic energy, whereas the two photons have an energy of 1 Mev $= 2mc^2$ and no mass. Here we have a striking example of a reaction where mass and kinetic energy are not conserved separately. But what we called the total energy in Section 32 - 2—that is, the sum of rest energy and kinetic energy—is conserved; the rest energy of the pair is $2mc^2$, equal exactly to the total energy of the photons, which is all kinetic.

32–7 NUCLEAR REACTIONS: CONSERVATION OF TOTAL ENERGY

The annihilation of the electron-positron pair is only one example of a reaction in which mass and kinetic energy are not conserved separately. Most reactions between atomic nuclei are of this kind. Let us look at some characteristic reactions.



32–11 When slow neutrons hit a block of paraffin, which contains many hydrogen atoms, 2.2 Mev gamma rays are observed. The gamma ray energy can be determined from the intensity of the light flash which is produced when the gamma ray interacts with electrons in the detector. (See Fig. 30–22.)

When neutrons from a reactor enter a block of material containing much hydrogen—paraffin, for example—they are slowed down by collisions with the nuclei of hydrogen atoms (protons) and finally combine with them to form the heavy isotope of hydrogen called deuterium. While this process goes on, gamma rays of 2.2 Mev are observed to emerge from the paraffin (Fig. 32 - 11). In this reaction, which we write as $n + H \rightarrow D + \gamma$, the neutron and proton are practically at rest. From where, then, comes the energy of the gamma ray? Let us look at the masses of the particles involved. The masses of the neutral atoms of hydrogen, deuterium, and some other isotopes which will be useful shortly are given in Table 2. From it we see that the sum of the masses of one neutron and one hydrogen atom is 2.0165 amu, whereas the mass of an atom of deuterium is only 2.0141 amu. Thus mass is not conserved in this reaction. The difference in the masses of 0.0024 amu corresponds to a difference of 2.2 Mev in the rest energy, as can be seen from the next-to-last column of the table. But this is just the energy of the emitted photon. Thus, as in the case of the electron-positron annihilation, the total energy, the sum of all rest energies and kinetic energies, is conserved.

TABLE 2

	MASS		REST ENERGY	
PARTICLE	IN AMU	IN KG ($\times 10^{-27}$)	IN MEV	IN JOULES ($\times 10^{-10}$)
n	1.00867	1.67481	939.55	1.50525
H	1.00782	1.67341	938.77	1.50399
D	2.01410	3.34426	1876.09	3.00568
T	3.01605	5.00791	2809.38	4.50090
He^3	3.01603	5.00787	2809.36	4.50086
He^4	4.00260	6.64600	3728.33	5.97315

All masses shown in the above table, except the mass of the neutron, were found by measuring the mass of positive ions of these atoms in a mass spectrograph. (See Chapter 28, Section 7.) The mass of the missing electron was then added to give the tabulated mass of the neutral atom. The neutron mass was determined indirectly from nuclear reactions.

1 amu $= 1.66042 \times 10^{-27}$ kg $= 931.476$ Mev and is 1/12 of the mass of the atom C^{12}.

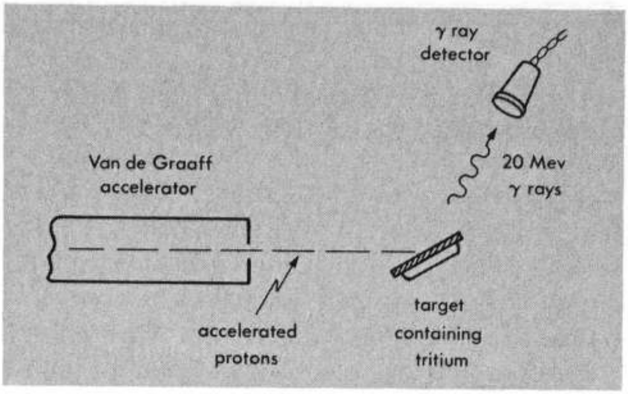
Here is another reaction, which does not involve neutrons: ordinary hydrogen combines with the heaviest isotope of hydrogen, called tritium, and yields helium; a very energetic gamma ray is produced in the process, $H + T \rightarrow He^4 + \gamma$. For this reaction to occur, the nuclei of the two hydrogen isotopes must get very close together. Since both are positively charged they repel each other. To overcome this repulsion the protons are accelerated to about 2 Mev and fired at a target containing tritium (Fig. 32 – 12). Again from Table 2 we see that $m_H + m_T = 4.0239$ amu, whereas $m_{He^4} = 4.0026$ amu, which corresponds to a difference in the rest energy of 19.8 Mev. The energy of the gamma ray is always around 20 Mev; it depends slightly on the direction in which the ray is emitted. This is not surprising, because the helium nucleus and the photon have to share the momentum of the incoming proton. When this is taken into account, the total energy balances exactly.

In the three reactions which we have discussed so far,

$$\begin{aligned} e^- + e^+ &\rightarrow 2\gamma \\ n + H &\rightarrow D + \gamma \\ \text{and } H + T &\rightarrow He^4 + \gamma \end{aligned}$$

the change in rest energy was taken up by photons. The next example of a nuclear reaction will demonstrate that rest energy can be converted into kinetic energy of other particles, not necessarily photons.

When a slow neutron is absorbed by a nucleus of U^{235} ("uranium 235") a new nucleus of U^{236} is formed. This nucleus has a high probability to split into two parts—for example, a nucleus of Ba^{142} and a nucleus of Kr^{94}. This process is called fission. Now, in reality the process is much more complicated than our brief description may indicate; before everything has settled down some neutrons, gamma rays, and electrons fly away from the fragments. But the main feature of the fission process is the formation of two fragments. A table similar to Table 2 will reveal that the rest energy of a neutron and an atom of U^{235}, that is, $(m_n + m_{U^{235}})c^2$ is larger than the sum of the rest energies of the fragments by about 200 Mev. Most of this energy is transformed into kinetic energy of the fragments, which fly off at high speed. It is this tremendous release of kinetic energy per atom which is at the heart of nuclear power.



32–12 An arrangement for detecting the reaction $H + T \rightarrow He^4 + \gamma$.

Let us review what we have learned in this section. We have seen three examples of reactions where an increase in kinetic energy was accompanied by a decrease in the rest energy of the particles involved. In Section 2 we introduced *total energy* and *rest energy* for the convenience of having names for the two terms which express the kinetic energy of a body of mass m

$$E_K = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2$$

Now we see that there is more to it than convenience. The rest energy of a body is not an abstract quantity of energy which is always locked up in the body and therefore can be regarded as an additive constant which might as well be ignored. Part of it—or all of it, as in the case of electron-positron annihilation—may change into another form of energy. And this is why the total energy becomes significant; as far as we know, it is always conserved.

32–8 MASS AND ENERGY

In our earlier work we developed the idea of mass as a measure for the quantity of matter. You will recall that we first found two seemingly independent measures, the inertial mass and the gravitational mass. Experiments have shown that the two kinds of mass are proportional to each other for all bodies and that we could make them equal by choosing the same unit for both. From then on we simply referred to mass without an adjective and used whichever method was most convenient. (See Chapter 13, Section 13 – 6.)

We have seen that mass is conserved in chemical reactions, provided, of course, the reaction takes place in a closed container and no molecules are allowed to escape. In the nuclear reactions which we discussed in the preceding section, mass is not conserved, although no atoms or electrons were allowed to escape. We notice also another difference. The energy changes which take place in chemical reactions, even in violent ones, are of the order of a few electron volts per molecule. The energy released per nucleus in a typical nuclear reaction is 10^6 times larger. Is there, perhaps, a connection between mass and energy which has escaped us as long as we were dealing only with small changes in energy?

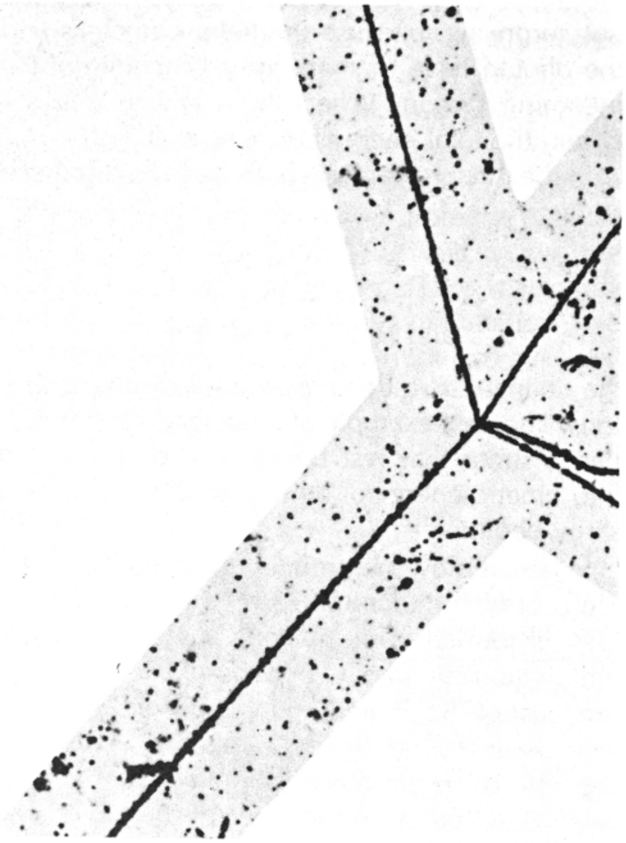
Consider again the formation of deuterium: $n + H \rightarrow D + \gamma$. We start with a neutron and a proton close together and at rest, and if we ignore the small recoil of the deuteron, we end up with a deuteron at rest and a photon. We saw already that the total energy of the system is conserved: $m_H c^2 + m_n c^2 = m_D c^2 + E_{photon}$. We can rewrite this equation as

$$m_H + m_n - m_D = \frac{E_{photon}}{c^2}$$

Rewritten in this way, the conservation of total energy says that the sum of the masses of the proton and neutron exceeds the mass of the resulting deuteron by the energy of the photon emitted in the process divided by c^2.

What model can we suggest by which a proton and a neutron form a deuteron, and yet combined have a total energy which is lower than their initial energy? Actually, we already know a suitable model. When an electron and a proton combine to form a hydrogen atom, one or more photons are emitted as the atom goes to its lowest energy state. In this ground state, the electron has both kinetic and potential energy. Their sum is negative, if we choose the potential energy to be zero at complete separation.

We can consider the formation of a deuteron in a similar way. While the deuteron as a whole is at rest, we can look upon it as a kind of double star, with the proton and the neutron orbiting around their common center of mass. (For more about the deuteron, see Section 35 – 11.) In forming the deuteron, the proton and the neutron gain kinetic energy, but this gain is more than offset by the reduction in potential energy due to the strong nuclear force. Initially the kinetic and potential energies are zero, but when the deuteron is formed their sum is negative. Thus we can express the overall energy of the proton and neutron in the deuteron as the sum of the individual energies (rest + kinetic) and their mutual negative potential energy. But this overall energy is just the rest energy of the deuteron $m_D c^2$. We see, therefore, that the mass of the deuteron is a measure for the overall energy of its constituents. If we considered a system containing any number of particles, we would



32–13 The absorption of a negative pion by a nucleus in a photographic emulsion as seen through a microscope. The pion came in from the right and was brought to rest, as indicated by its meandering path. The mass of the nucleus has increased by the mass of the pion, which is 0.151 amu. The resulting increase in energy, 141 Mev., is largely in the form of kinetic energy. At least four high-speed particles, characterized by their straight paths, can be seen; there may be additional neutral fragments which do not leave tracks in the emulsion. (From: Powell, Fowler, and Perkins The Study of Elementary Particles by the Photographic Method, Pergamon Press.)

arrive at the same conclusion, namely that the mass of the system is a measure for the overall energy—rest, kinetic, and potential energy of its constituents. If the energy changes, the mass changes accordingly and the two changes are related by $\Delta m = \Delta E/c^2$.

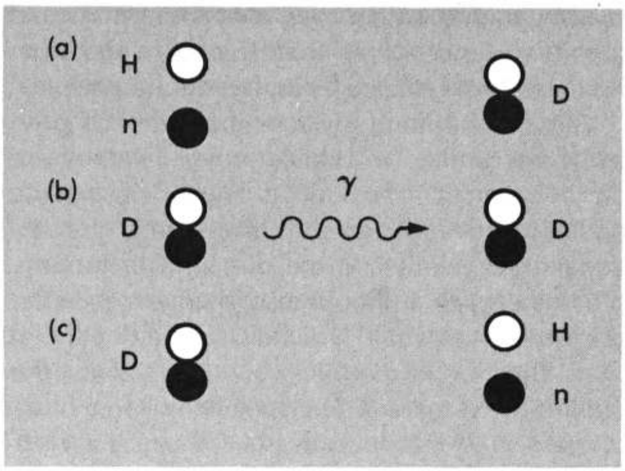
Notice the value of the scaling factor $1/c^2$ in $\Delta m = \Delta E/c^2$; in standard units it is about 10^{-17} (sec/m)^2. Even a huge addition of energy makes for only a minute addition in mass. Conversely, rewriting the equation as $\Delta E = (\Delta m)c^2$ tells us that a tiny change in mass results in a tremendous change in energy. When this energy is in the form of kinetic energy, it is very apparent indeed (Fig. 32 – 13).

We now see that our early definition of mass as a measure for the quantity of matter is useful only as long as we restrict ourselves to processes accompanied by small changes of energy. In the formation of the deuteron or in other nuclear reactions the number of protons and neutrons does not change, so in a sense the quantity of matter remains the same. Nevertheless, we have seen that the masses of the nuclei involved have indeed changed.

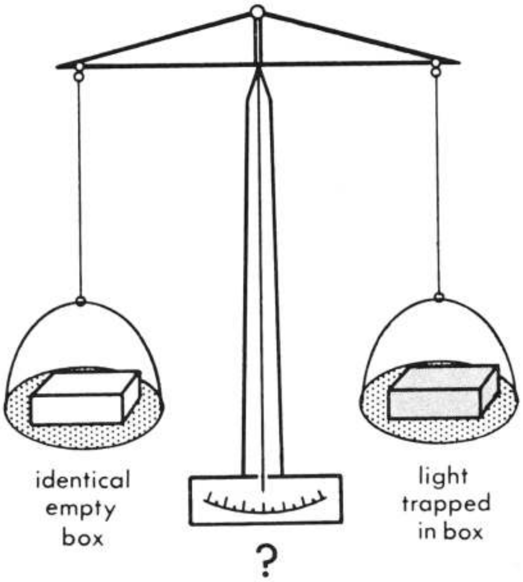
32–9 MASS FOR PHOTONS?

Suppose now that the photon emitted in forming the deuteron is absorbed by another deuteron at rest nearby, which then disintegrates into a proton and a neutron (Fig. 32 – 14). If we again ignore the recoil energy, we have the same collection of particles at rest before and after the emission of the photon. Their total mass is unchanged. The only result of the emission and absorption of the photon is a change in the distribution of mass. In Fig. 32 – 14(a) the proton and neutron are on the left and the deuteron on the right, whereas in Fig. 32 – 14(c) it is the opposite. An amount of mass equal to E_{photon}/c^2 was carried by the photon from one place to another. It is tempting to ask whether the photon only transfers mass by being emitted at one place and absorbed at another, or whether it possesses by itself mass equal to E_{photon}/c^2.

To answer this question we must have a way to measure the mass of a photon. One thing is certain, we cannot hold a photon at rest and put it on a balance; photons always move with the speed c. We might perhaps think of the following experiment. Consider two identical boxes with completely reflecting walls (Fig. 32 – 15). One has nothing in it, and in the other we trap a large amount of light. We can put both boxes on an equal-arm balance and compare their masses. If we had perfectly reflecting walls and a sensitive enough balance, this experiment would answer the question. But you can easily convince yourself that it cannot be done in practice. To decide whether the photon itself has a



32–14 Schematic description of the three stages in the creation of a deuteron from a proton and neutron, emission of a gamma ray, and the subsequent dissociation of a second deuteron.



32–15 The two boxes are in balance when empty. What will happen if a large quantity of light is trapped in one of them?

mass, we must extend our concept of mass and admit evidence other than that obtained from an equal-arm balance or an inertial balance.

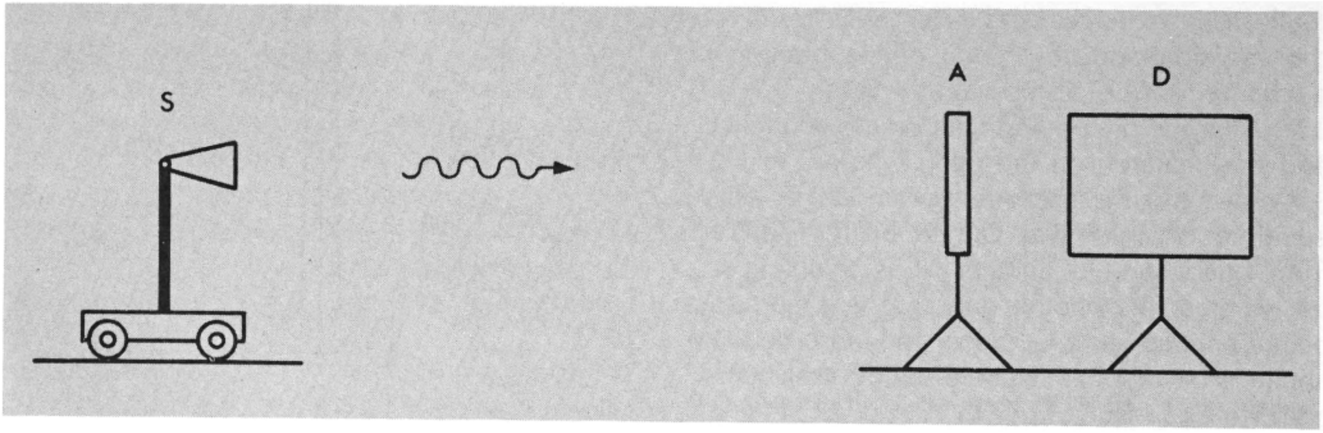
This can be done because the action of gravity is not limited to bodies at rest (as they are on a balance). The kinetic energy of moving bodies changes when they move in a gravitational field parallel to the direction of motion. If the energy of a photon also changes when the photon moves in a gravitational field parallel to its direction of motion, we would say that the photon has a mass. To be specific, when a body of mass m is thrown upward a distance d near the surface of the earth, its kinetic energy decreases by $\Delta E = mgd$. We could say that a photon of energy E has a mass E/c^2 if by moving up a distance d its energy decreases by $\Delta E = (E/c^2)gd$.

Let us see what the size of this effect might be. The fractional change in the energy is $\Delta E/E = gd/c^2$. A reasonable height inside a building is about $d \approx 10$ m; then with $g \approx 10$ m/sec^2 and $c = 3 \times 10^8$ m/sec we find $gd/c^2 \approx 10^{-15}$. To decide whether a photon has a mass, in the extended sense of the word, we must be able to detect a change in its energy of the order of one part in 10^{15}! It almost looks as if this experiment must remain a thought experiment, like weighing a box with light trapped in it.

There is one difference, however: the energy of a photon is proportional to its frequency $E = h\nu$ (where h is Planck's constant) and therefore the change in energy is proportional to the change in frequency: $\Delta E/E = \Delta\nu/\nu$ and changes in frequency can nowadays be measured with extremely high accuracy. The change of energy of gamma rays with height was actually observed in 1960 by R. V. Pound and G. A. Rebka, Jr., in a tower about 25 m high at Harvard University.

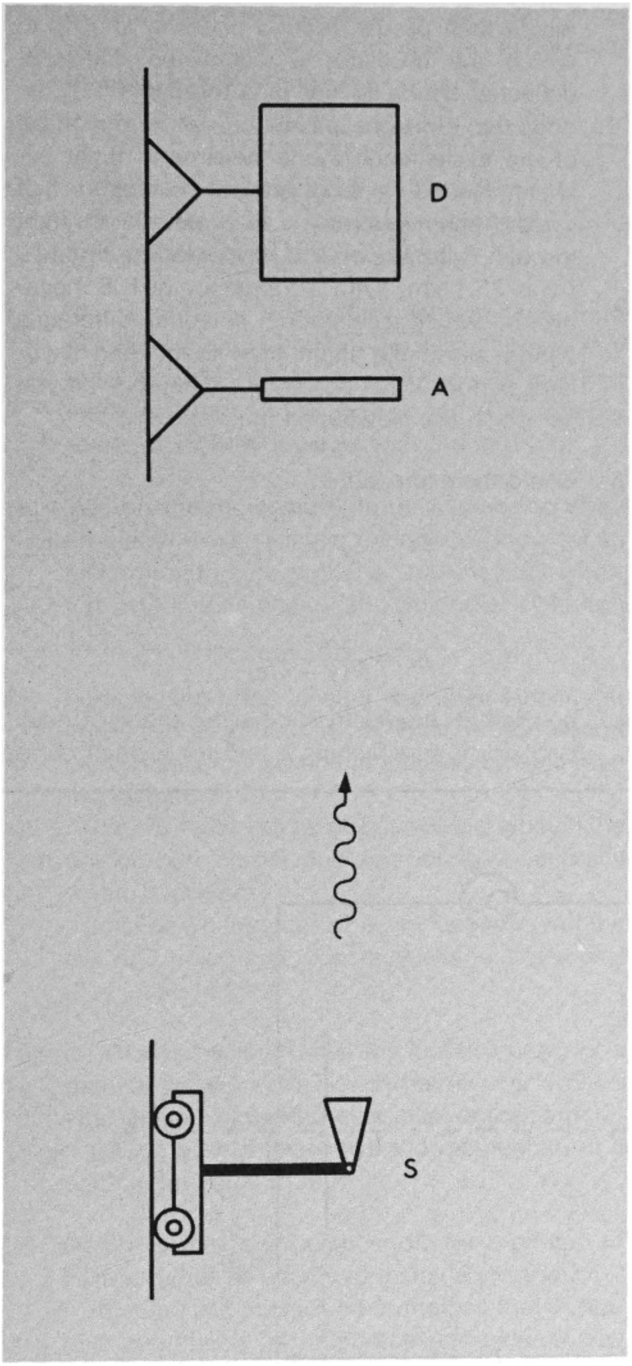
The technical details of Pound and Rebka's experiment are extremely complex, but the principle is quite simple. Consider a source of monochromatic light (*S*) and an absorber (*A*) which strongly absorbs light with exactly the frequency of that emitted by the source, and is quite transparent to other light even if its frequency is only slightly different. Behind the absorber there is a detector (*D*) which detects the light which passes through the absorber (Fig. 32 – 16). When *S* is at rest relative to *A*, the absorption is largest. When *S* is moving toward *A*, the frequency of the photons will increase ("blue shifted"; see Appendix to Chapter 30), fewer will be absorbed by *A*, and therefore more will reach *D*.

Now suppose we arrange the source, absorber, and detector along a vertical instead of a horizontal line (Fig. 32 – 17) and consider first the case when *S* is at rest with respect to *A*. If the frequency of the photons on the way up decreases owing to the gravitational pull—that is, if there is a gravitational "red shift"—then fewer photons will be absorbed at *A* than in the horizontal arrangement (Fig. 32 – 16). On the other hand, by moving the source upward, we can give the photons an initial "blue shift" to compensate for the gravitational "red shift" so



32–16 Schematic drawing of the main components of the Pound and Rebka experiment. In the horizontal arrangement shown, maximum absorption occurs when the source S and the absorber A are at rest with respect to each other.

that they arrive at A with just the right frequency to be strongly absorbed. We can express this same idea in terms of energy: by emitting the photons from a source moving upward, we give them a little additional energy to compensate for the loss of energy they experience on the way up. Thus, if photons have a mass, the strongest absorption should occur when the source is moving upward at just the right speed to cancel the gravitational "red shift." This is the essence of Pound and Rebka's experiment. Their results, and results of later experiments, confirmed to within 1 percent that photons indeed behave as though they have a mass of $m = E/c^2$.



32–17 **In the vertical arrangement maximum absorption occurs when the source is moving upward extremely slowly.**

The gravitational "red shift" was predicted by Albert Einstein long before the nuclear reactions discussed in Section 32 – 7 were discovered. The first observations of a gravitational "red shift" were made on light coming from very heavy and compact stars (white dwarfs). But the effect was not as direct and convincing as that observed by Pound and Rebka.

32–10 REVIEW AND OUTLOOK

Let us review briefly the ground we have covered in our study of relativity. We started with an unanswered question in optics: in what medium do light waves propagate? Our search did not really answer the question; instead it resulted in a revamping of the fundamental ideas of kinematics. To be sure, Galilean kinematics was not proven to be wrong; we found its limitations and learned how to extend it. The chapter on relativistic kinematics ended with another question: how to reconcile the universal speed of light with Newtonian dynamics, which recognizes no "universal speed limit" but predicts that a body can be given any speed.

In this chapter we found that Newtonian dynamics also has its limitations, and again we learned how to modify the theory and extend it over a wider range. In the process we learned that the two conservation laws of mass and energy are actually one law.

Throughout the development of physics the study of light always had surprises in store for us. It has now had another. We first showed that photons had no mass in the classical sense of the word, only to find that they do have a mass in a more general sense. Looking back, you may properly be impressed by the long road we have covered. But there is even more ahead. Let us end this chapter by raising one more question. When we shine light upward, its frequency changes. But frequency is a purely kinematical quantity of dimension 1/time. Does time and perhaps length depend on the gravitational field? Do they change in the vicinity of heavy bodies?

QUESTIONS

1.* In Fig. 32 – 1 the distance between *A* and *B* was 8.4 meters. What was the speed of the electrons which caused the pulses shown in Fig. 32 – 2?

2.* Compare the kinetic energy of a particle of mass *m* and speed 0.600 *c*, as calculated relativistically, with the kinetic energy as calculated classically.

3.* How would you change the energy scale in Fig. 32 – 4 so that the graph would represent E_K vs. v/c for a proton?

4. What is the order of magnitude of energy required to accelerate a 1-kg mass from rest to a speed of 0.600 *c*? Compare it with the energy released by a 1-megaton hydrogen bomb. (A 1-megaton bomb is equivalent to 10^6 tons of TNT; TNT liberates about 10^7 joules per kilogram.)

5. What speeds would be acquired by (a) electrons and (b) protons which move from rest through potential differences of 10^5, 10^7, and 10^9 volts?

6. The mesons described in Section 30 – 15 were observed to have a mean lifetime greater than their rest lifetime by a factor of 9. What kinetic energy would you expect to measure for these mesons? The rest energy of a mu-meson is 106 Mev.

7. An early time-of-flight experiment with electrons used a radio-frequency oscillator, as in Fig. 32 – 18. Electrons from a hot cathode, *K*, are accelerated in a vacuum through a potential difference, *V*. Slits S_1 and S_2 limit the beam, which then passes through plates P_1 and P_2, to which the oscillator is connected. Electrons deflected by P_1 do not pass through slit S_3, so only the electrons passing P_1 at zero voltage of the cycle reach P_2. If the time of flight between P_1 and P_2 is an integral number of half cycles, the electrons will continue straight through P_2 to a reference spot at *R*. The distance *l* was 75.1 cm. With a frequency of 108 megacycles, what accelerating potential difference would cause the flight time to be one oscillator period? What percentage difference do you get if you use Newtonian mechanics?

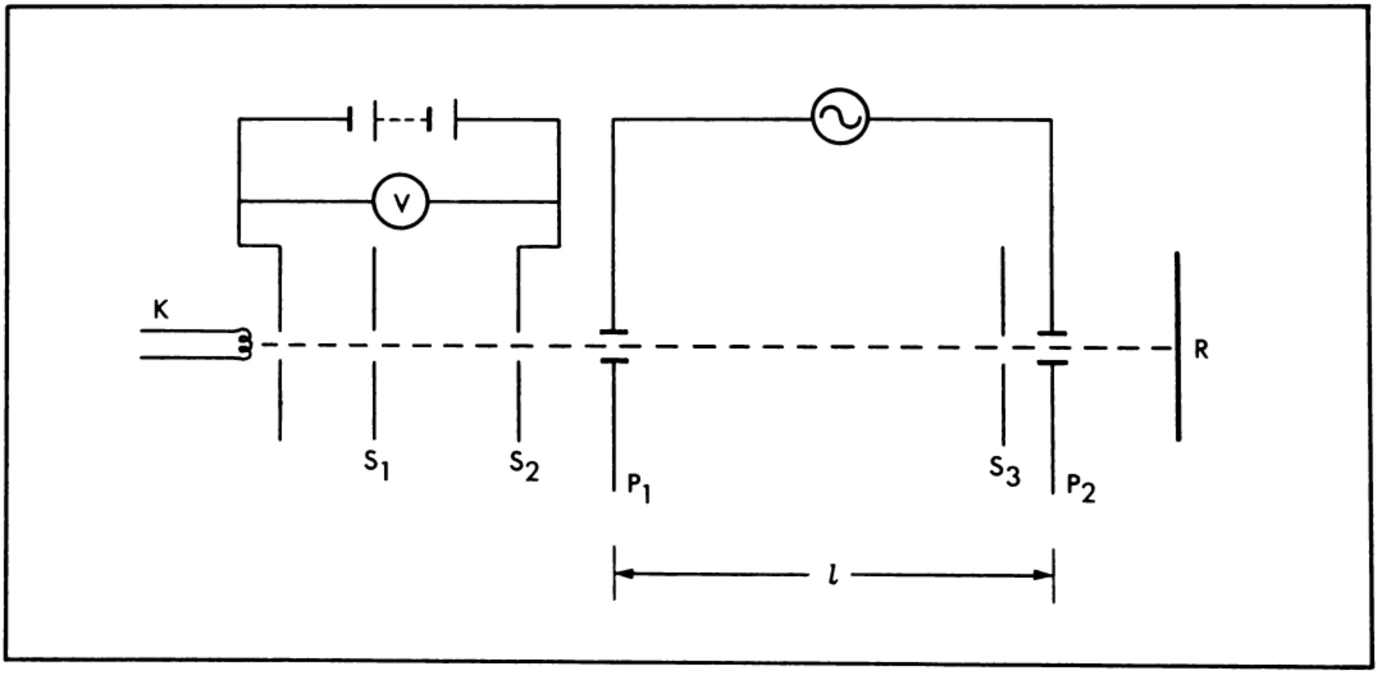
8 Could the expressions

$$E_K = \frac{\frac{1}{2}mv^2}{\sqrt{1 - v^2/c^2}}$$

and

$$E_K = \frac{mv^2}{2(1 - v/c)}$$

be used for the relativistic kinetic energy? Check each for *v* approaching *c* and for *v* small com-



32–18 For Problem 7.

pared to c. Does a change in sign of v make any difference? Also try each with a numerical value such as $v = 0.600\ c$.

9.* Compare the momentum of a particle of mass m and speed $0.600\ c$, as calculated relativistically, with the momentum as calculated classically.

10.* If a 1-kg mass is traveling at 1 meter/sec, what is its momentum in Mev/c?

11.* *Estimate* the momentum of an electron which has been accelerated from rest by a potential difference of a billion volts.

12. (a) From $E = \frac{mc^2}{\sqrt{1 - v^2/c^2}}$ and $p = \frac{mv}{\sqrt{1 - v^2/c^2}}$, eliminate v and express E as a function of p.

(b) For $p \ll mc$, expand the expression you find to the second order in p/mc.

13. Plot a graph of momentum versus energy from the expression $p = \sqrt{E^2 - (mc^2)^2}/c$. Use energy units of mc^2 and momentum units of mc. Plot the graph with E up to 500 mc^2. On a larger scale, show how p varies with E up to 5 mc^2.

14. What is the momentum, in units of Mev/c, of (a) an electron with kinetic energy 2.0 Mev, (b) an electron with kinetic energy 200 Mev, (c) a proton with kinetic energy 200 Mev? (See Question 13.)

15. Particles are often labeled with their kinetic energy—for example, a "10-Mev proton"—and calculations of momentum and speed made from that datum.

(a) For what values of E_K/mc^2 and v/c will the use of $E_K = \frac{1}{2}mv^2$ and $p = mv$ give accuracy within 1 percent?

(b) For what values of E_K/mc^2 and v/c will the use of $E = pc$ give accuracy within 1 percent? (E = total energy.)

16. (a) At what velocity will the kinetic energy of a particle be equal to its rest energy, mc^2? Does your answer depend on the mass of the particle?

(b) At what velocity will the momentum of a particle be equal to mc?

17. (a) Show that it is not possible for a photon to transfer all of its energy to a single electron.

(b) Can you suggest why photons may disappear completely in the photoelectric effect? (See Section 31 – 2.)

18. The intensity of the sun's radiation just outside of the earth's atmosphere is about 8×10^4 joules per square meter per minute.

(a) What pressure would this exert on a reflecting satellite like the Echo balloon?

(b) If we assume that the earth completely absorbs the sun's radiation, what would be the total force on the earth from this cause? How does this compare with the sun's gravitational force on the earth?

19. (a) How much energy is carried by an "average" photon of visible light with wavelength of about 5000 angstroms? How much momentum?

(b) Estimate the number of photons of visible light emitted per second from a 100-watt light bulb emitting 1 percent of its power in the visible region.

(c) What is the light pressure exerted by the photons from that bulb when they strike a black body 2 meters away head on?

20. When an atom emits a photon we usually equate the photon energy to the change in internal energy of the atom, $\Delta E = h\nu$. Actually, a part of the internal energy goes into recoil kinetic energy E_K of the emitting atom, so that $h\nu$ is really less than ΔE. Why? Calculate the ratio $E_K/h\nu$ for the yellow line of sodium, wavelength 5890 Å. The mass of the sodium atom is 3.8×10^{-26} kg.

21. In the film "Pressure of Light," it was stated that the strong light source used in the experiment exerted a pressure of about 10^{-9} atmosphere on the highly reflecting metal foil of the torsion pendulum. (1 atmosphere $\approx 10^5$ newtons/m^2.)

(a) What must have been the energy per unit volume of that light?

(b) About how many photons per cubic centimeter would that represent, taking 5500Å as the average photon wavelength?

22. A plane mirror is illuminated perpendicularly to its surface by light of frequency ν. The surface of the mirror is hit by N photons per unit time and per unit area.

(a) What is the change in momentum of a single photon when it is reflected by the mirror?

(b) What is the pressure of light on this mirror?

(c) Express the pressure in terms of the total energy E falling on the mirror per unit time and per unit area. Would it make any difference if the light, instead of being monochromatic, were spread over a whole range of frequencies?

(d) What is the numerical value of this pressure, assuming that the intensity of illumination is 10^3 joules/m^2-sec? (This intensity corresponds roughly to direct sunlight illumination at noon in summertime.)

23. In the Compton scattering experiment described in Section 32 – 5:

(a) What was the wavelength of the incoming photon?

(b) What was the fractional change in wavelength?

24. A 50,000 electron-volt X-ray photon makes a direct impact with an electron at rest and is scattered through 180°.

(a) What energy will the photon lose?

(b) What will be its change in wavelength? What fractional change is this?

25. A photon of wavelength $\lambda = 1.00$ Å hits an electron "at rest." The photon bounces away at right angles to its original path.

(a) Draw a vector diagram of the photon momentum before the collision and the photon and electron momentum after the collision.

(*Note:* For such a collision the fractional loss of energy is small; ν' is nearly equal to ν, so in drawing a vector diagram for momenta you can take the momentum of the photon after collision to have practically the same magnitude as the momentum of the photon before collision when you are finding the direction of the electron's recoil. Taking the electron recoil given by that assumption, use conservation of momentum and conservation of energy to compute the answers asked for.)

(b) Calculate the momentum of the photon after the collision.

(c) Calculate the kinetic energy of the electron after the collision.

(d) Calculate the energy of the photon after the collision.

(e) Is ν' very different from ν?

26. Suppose the electron in Question 24 is moving toward the X-ray photon with a momentum which is equal in magnitude to that of the photon. If the photon is again scattered through 180 degrees, (a) what energy transfer would take place? (b) What change in photon wavelength would occur?

27.* An electron and positron, both essentially at rest, annihilate each other. Find the frequency and wavelength of the photons emitted.

28. An electron at rest and a positron with a kinetic energy of 10 ev create two photons by annihilation. What will be the angle between the photons (a) if one moves in the direction the positron had? (b) If one moves at right angles to the direction the positron had? How will the energies of the two photons compare in each case?

29. Consider qualitatively a momentum vs. energy graph for electrons or positrons (see Question 13) and the graph of $p = E/c$ for photons. Why is it not possible for a single photon to produce an electron-positron pair spontaneously in empty space? Can you suggest the circumstances under which pair creation occurs?

30. What gamma-ray energy would you expect from the reaction $\mathrm{n} + \mathrm{He}^3 \rightarrow \mathrm{He}^4 + \gamma$ if the incident slow neutron has negligible kinetic energy?

31. A neutron and a proton at rest form a deuteron, and a photon is emitted. Calculate the recoil kinetic energy of the deuteron E_K and the ratio $E_K/h\nu$, where $h\nu$ is the energy of the photon. Compare the result with what you found in Question 20.

32. One way to obtain fairly high-energy neutrons is to bombard a target containing tritium with deuterons of about 5 Mev. The reaction is $\mathrm{D} + \mathrm{T} \rightarrow \mathrm{He}^4 + \mathrm{n}$.

(a) Show by means of a diagram what determines the relative directions in which the products are emitted.

(b) In which direction would the most energetic neutrons be emitted?

33. A reaction rather like that of Question 32 occurs when tritium is bombarded by protons of suitable energy: $\mathrm{H} + \mathrm{T} \rightarrow \mathrm{He}^3 + \mathrm{n}$. Confirm from Table 2 that the rest energy of the products exceeds that of the reacting particles by 0.8 Mev. Why can you be sure that the reaction would not occur if the incident proton had a kinetic energy of exactly 0.8 Mev?

34. The nuclear reaction $\mathrm{D} + \mathrm{D} \rightarrow \mathrm{H} + \mathrm{T}$ occurs when 2.4 Mev deuterons strike a thin deuterium target.

(a) What is the sum of the kinetic energies of the products?

(b) What determines the way in which this sum is shared?

35. We may look at the fission of a U^{236} nucleus into two nuclei of Ba^{142} and Kr^{94} as taking place in two stages. The protons and neutrons first rearrange themselves in two separate groups and then fly apart. If this picture is correct, then the final kinetic energy of the two fragments must equal the electrostatic potential energy of re-

pulsion between the two fragment nuclei just before they fly apart. The latter will depend on the initial distance between the two nuclei just after they have formed up but before they fly apart. Calculate the order of magnitude of this distance, and check your answer with what you know about nuclear radii from Chapter 26.

36. Taking recoil into account, the emission and reabsorption process depicted in Fig. 32 – 14 is really not possible. Discuss qualitatively how you might modify the setup slightly to make it possible.

37. The sun radiates light of all frequencies at the approximate rate of 3×10^{26} joules per second.
    (a) What is the mass lost by the sun per second? In 5 billion years (the approximate age of the solar system)?
    (b) How does this affect the earth's orbit?

38. Calculate the relative change in mass of a hydrogen atom when it makes a transition from its first excited state to its ground state.

39. It takes 3.4×10^5 joules/kg to melt ice. What is the relative increase in mass when an ice cube melts?

40. Suppose you have two cubes of steel, each weighing 10.0000 grams, in balance on an analytic equal-arm balance when both cubes have the same temperature. You want to find out whether the same cube heated up has a larger mass, so you heat one cube and cool the other and put them back on the balance. Discuss some of the technical problems you will have to solve before you can expect to get a meaningful answer from your experiment.

41.* (a) If the source were above the absorber and detector in Fig. 32 – 17, which way would the photon frequency be shifted by gravity?
    (b) In which direction would the source have to move to compensate for this?

42. Assume that all the light used in New York City in one hour in the evening is trapped in a box. How much will the mass of the box increase? Estimate, as best you can, the power consumption for light.

43. Suppose you could do an experiment analogous to that of Pound and Rebka with light of wavelength 6000Å using a tower 100 m high. What must the speed of the source be to cancel the gravitational red shift?

44. The gravitational red shift was first measured in the light received from certain very dense stars called white dwarfs. They have a mass about equal to that of the sun and a radius only as large as that of the earth. What is the fractional shift in energy of a given line in the spectrum of a white dwarf compared with the same line produced in the laboratory?

45. In a sodium-vapor lamp, emitting light of wavelength 5890Å, the sodium atoms have kinetic energies associated with a temperature of about 500°C. Calculate the average speed of sodium atoms with this thermal energy, and use it to find the difference in wavelength of the observed light emitted by atoms approaching and receding from the observer. (Recall that $E = h\nu = hc/\lambda$, and see the Appendix to Chapter 30.)

46. The mass and radius of the sun (M,R) and of the earth (m,r) and the mean distance between them (L) are given in Table 2 of Chapter 15. Consider a particular spectral line emitted by the sun.
    (a) Is this line shifted to the red or to the blue end of the spectrum, compared with the same spectral line when produced in the laboratory? Calculate the fractional shift.
    (b) If the hot gases on the surface of the sun, from which this spectral line comes, are not at rest but streaming radially outward, then the light emitted from the central part of the sun's disc actually comes from atoms moving toward us. What must the streaming velocity v be in order to cancel the gravitational shift?
    (c) Can a red shift due to motion of the emitting atom be distinguished in general from a gravitational red shift?
    (d) Can the Doppler shift in the light from the sun be distinguished from the gravitational shift?

47. Does the light bend when it passes near a heavy body like the sun? Explain your answer on the basis of what you have learned in this chapter.

33

Atoms and Spectra

IF WE LOOK back over the wide range of phenomena which can be summarized in terms of Newtonian mechanics, we find the record truly impressive. The conservation laws of momentum, energy, and angular momentum span the field from the motion of the planets down to the motion of molecules in a gas. The inverse-square law for the interaction between masses or electric charges governs the motion of a rocket to the moon as well as the motion of an alpha particle passing near a gold nucleus. When we tried to apply Newtonian mechanics to particles moving at speeds close or equal to the speed of light we encountered some problems, but it turned out that the conservation laws hold also in this domain and that only the relations between kinetic energy, momentum, and velocity had to be revised.

Let us now try to pull together all our knowledge of mechanics and electromagnetic waves and apply it to individual atoms. That is, let us see what happens when we try to apply the laws of mechanics and electromagnetism to objects whose dimensions are 10^{-10} m. We shall do this using the Rutherford model described in Chapter 26.

33–1 THE STABILITY OF ATOMS

The Rutherford model, describing the atom as a tiny but massive positively charged nucleus surrounded by a number of electrons, seems essentially correct. Moreover, we have learned how to measure the charge on the nucleus and how to estimate at least an upper limit for its "radius." Knowing the charge of the nucleus, we also know the number of electrons which complete the neutral atom.

And yet, as we continue our study, our complacency begins to be shaken. A serious flaw in the Rutherford model becomes evident as soon as we start considering the behavior of the electrons. In the atom, as imagined by Rutherford, the electrons move around the nucleus like planets around the sun. In fact, under the urging of the Coulomb force, they can avoid falling immediately into the nucleus only by continuously revolving around it. This motion is an accelerated motion, and therefore should be accompanied by the emission of electromagnetic waves. But the electromagnetic waves drain away energy. So, like an earth satellite which loses energy in the atmosphere, the electrons must spiral into the nucleus. By imagining that they move like planets, we have only stretched the time for electrons to fall in. Instead of falling into the nucleus in about 10^{-17} sec, as an electron would if it started from rest at the outside of an atom, an electron spiraling in while radiating away its energy should take about 2×10^{-11} sec to get to the nucleus. This time is still far too short to make the atom look stable at the dimensions we know. Consequently, Rutherford's model is not consistent with our knowledge of the process of electromagnetic radiation.

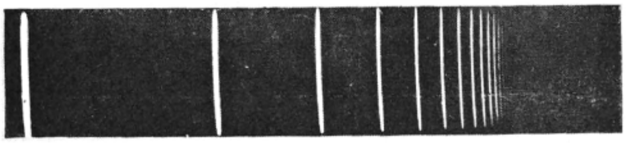
Atoms do emit light, and that light must come from the accelerated motion of the charge in the atom. Therefore our whole picture of the atom is clouded by the conflicting demands. The emission of light is consistent with the Rutherford model—but it also destroys that model, for the atoms should all shrink to nuclear dimensions long before we can blink at their light. The normal atom should have the electrons stuck in the nucleus into which they have fallen. But this is in direct conflict with the evidence of alpha-particle scattering and with the size of atoms as seen in collisions in gases and in their packing in solids and liquids.

The more we probe into these questions, the more we become aware of serious troubles. Let us consider for simplicity the hydrogen atom, consisting of a single proton and a single electron. The frequency of the light emitted by the atom should be related to the number of revolutions per second of the electron around the nucleus. But the period of revolution depends on the diameter of the orbit, just as in the case of the planets in the solar system (Chapter 15). Electrons rotating in a smaller orbit have a shorter period and therefore should emit light of a higher frequency than electrons rotating in larger orbits.

As an electron emits light, its energy must decrease. Therefore, the diameter of its orbit must decrease and the frequency of the emitted wave must increase. A light source contains a huge number of atoms, and many are emitting light at one time. Some of these atoms are at one stage in the process of light emission, and some at another. Consequently, the source should be emitting light of practically all frequencies. Thus, hydrogen gas made to glow, for example, by an electric discharge should emit a continuous spectrum of light. In contrast to this prediction, spectral analysis of the hydrogen light reveals a number of sharp "lines" – that is, a number of separate wavelengths or frequencies (Fig. 33-1).

Apart from this basic trouble—energy loss by radiation—there is another reason to fear that the Newtonian laws of mechanics can never explain the Rutherford atom. We can picture some agent (such as a visiting star) changing our own solar system by any amount, small, medium, or large—short of a major breakup—and when the disturbing agent was removed the new arrangement would just continue. We can imagine a solar system of planets built to any size over a wide range—with orbits, masses, and numbers of planets different from ours, and yet equally stable. A collection of similar solar systems, such as we might find around other stars, would be like the other things we see in nature, like a set of roses or of raindrops; they would bear a strong family resemblance, but with different-sized suns, and planetary orbits of different sizes, no two would be precisely alike.

Atoms are not like that. Their most remarkable property is the likeness they bear one to another. All hydrogen atoms act the same, all helium atoms are alike, and so on. Atoms of one kind are far more alike than any Grade A peas. Yet they have different histories; different stories of collisions and ionizations and chemical combination lie behind each one. Through it all they remain one and the same, showing the same spectral lines and making the same chemical reactions with the same amounts of energy. The atom may be a planetary system, but it is a planetary system that comes in myriad identical copies.



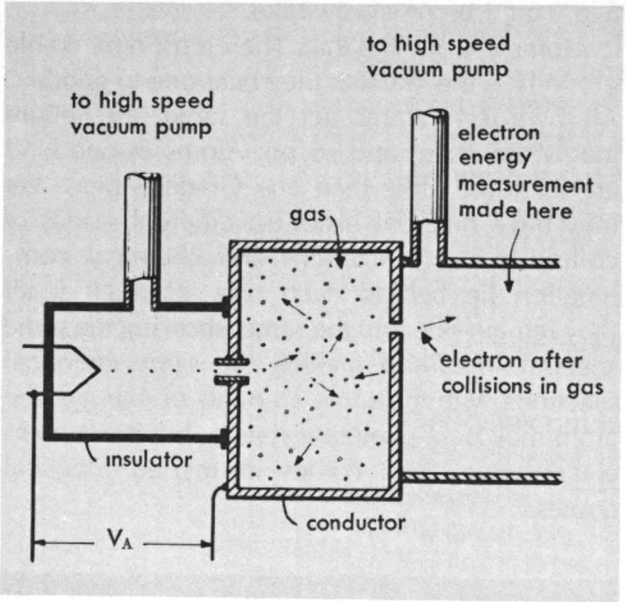
33-1 A photograph of the radiation from glowing hydrogen gas, taken through a spectroscope. Instead of a continuous spectrum of wavelengths, only certain discrete wavelengths are produced.

33–2 THE EXPERIMENTS OF FRANCK AND HERTZ; ATOMIC ENERGY LEVELS

For a planetary system, according to Newtonian mechanics, motions at all total energies are possible. For an atom, on the other hand, apparently only a limited set of motions can take place. This limitation is suggested by the repetitious behavior and by the stability of atoms. The identity of the members of an atomic species suggests the same thing.

If only certain distinct patterns of motion are possible, we would expect that an atomic system can take on only the separate total energies corresponding to those motions. Then there should be gaps between the possible energies; and in that case the internal energy of an atom can be changed only by certain definite amounts of energy transferred from the outside.

Are there really gaps between the possible energies that an atom may have? A direct test of this idea is important. It can be made by trying to change the energy of atoms directly by bombarding them with electrons. Can an atom that is hit by an electron take in some energy and keep it as a gain of internal energy? And if so, can it retain *any* gain from small to large, or is it limited to a selection of sharply defined amounts? In Germany in 1914 such a test was made. James Franck and Gustav Hertz performed an electron-bombardment experiment for which they later received the Nobel Prize. This experiment—and the many similar ones that stem from it—are a key to the problems of atomic stability and structure. Here we shall describe an experiment somewhat different from the original Franck and Hertz experiments which gets the same information by closely related means. We choose this particular experiment because it is easier to interpret though harder to carry out.


to high speed vacuum pump
to high speed vacuum pump
electron energy measurement made here
gas
electron after collisions in gas
insulator
conductor
V_A

33–2 The general idea of an experiment to measure the changes in the energy of electrons when they collide with gas atoms. The electrons pass through a sample of gas (mercury vapor) in the middle chamber. Electrons leave the gun with energy given by the accelerating voltage. Their energy remaining after collisions is measured in the right-hand chamber. Note that the opening into the energy-measuring region is not in the direct line of the original beam of electrons. This is to insure catching only those electrons that have been deflected by collisions.

Like Franck and Hertz we shall use electrons to bombard the atoms. We can do this by using an electron gun like the ones we described in Chapter 23. The electron gun gives kinetic energy to the electrons boiled out of the filament. If the accelerating voltage in the gun is V_A, the electrons leave the gun with kinetic energy of one electronic charge times V_A.

When the electrons leave the gun they pass into a chamber containing a gas. They enter this chamber through a small hole, and the electron gun operates in a good vacuum because we continually and rapidly pump away any gas that escapes from the chamber into the gun (Fig. 33-2). The chamber containing the gas is made of a conducting material so it is all at one potential and there is no electric field inside it which could change the energy of the electrons. Consequently, any change in the electrons' energy arises their interaction with the gas atoms.

Now to investigate what happens to the electrons when they collide with the gas atoms, we allow some of the electrons to come out of the gas chamber through another small hole leading out on a slant into a second evacuated space. There we measure the energy of the electrons which have passed through the gas. We can make this measurement with any one of a number of devices. For example, we can use a magnetic mass spectrograph like that described in Chapter 30. Then, since we know the mass of the electron, the curvature of the electron path in the magnetic field allows us to determine the energy. Alternatively, we can determine the electron's energy by making the electron move against a retarding potential difference and measuring what potential difference just prevents the electrons from reaching the collector. We could even use time-of-flight measurements in the detecting chamber to find electron speed and therefore kinetic energy. The practical details of these measurements turn out to be difficult and complicated, but despite the complications we emerge with the knowledge we want of the energy of electrons after interaction with the gas atoms.

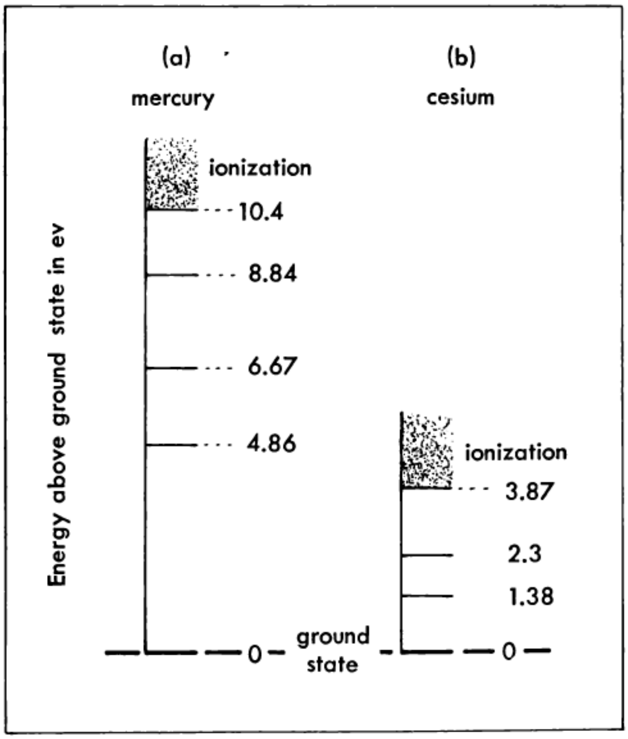
What kind of information do we get from this apparatus? Let us take a specific example. What happens when we have mercury vapor in the gas chamber? As long as the accelerating potential in the electron gun is only a few volts, we find that the electrons entering our detector after colliding with the gas atoms have almost exactly the same energy that is imparted to them by the gun. The collisions with gas atoms at this low energy are elastic. (In such collisions the massive atoms take a negligible share of kinetic energy.) However, when the accelerating potential in the gun passes 5 volts, so that the electrons bombard the atoms with 5 electron volts or more kinetic energy, a dramatic change occurs. The electrons entering the detector no longer have the energy with which they left the gun. Instead, the electrons entering the detector have practically no kinetic energy.

Now we increase the gun potential further so that the kinetic energy of the bombarding electrons is greater. We find that the energy of electrons reaching the detector is also greater. When we raise V_A from 5 volts to 6 volts, the kinetic energy measured in the detector rises from a small fraction of an electron volt to one electron volt more than that. Careful measurements in which we determine the gun voltage accurately and the kinetic energy in the detector accurately show that in their collisions with the atoms the electrons are losing almost exactly 4.9 electron volts. This loss only sets in when the bombarding electrons have 4.9 electron volts of kinetic energy, and the amount of the loss stays the same while we increase the bombarding energy through the next couple of electron volts. The mercury atom cannot accept anything less than 4.9 electron volts of energy, and even when it is offered somewhat more energy it still accepts exactly this amount.

However, 4.9 electron volts is not the only amount of energy the atom can take in. If the bombarding energy is 6.7 electron volts or more, the electrons can lose either 4.9 electron volts or 6.7 electron volts. On increasing the bombarding energy still further, we find other thresholds at which greater losses can occur and beyond which any of several different amounts of energy can be robbed from the electrons. This is the essential property Franck and Hertz looked for: atoms can change their internal energy, but the changes are restricted to sharply defined steps.

The smallest amount of energy which can be accepted by an atom is called its first *excitation energy*. For mercury the excitation energy is 4.9 electron volts. For helium, which has the highest excitation energy, the value is 19.8 electron volts. The lowest excitation energy of all the atoms goes to cesium, which will accept a bundle of 1.4 electron volts. Other atoms will also accept certain definite bundles of energy, and these bundles are characteristic of the kind of atom.

The experiment we have described is an idealization; the results we have quoted were collected in a number of similar experiments, all of which agree. Thus the Franck-Hertz experiment and its later modifications have shown us that atoms can accept only parcels of energy. The internal energy content of an atom cannot be continuously changed. It changes only in steps. The successive internal energies which an atom can take on are called its *energy levels*. Often they are diagrammed as in Fig. 33 – 3.



33–3 (a) Some of the energy levels of the mercury atom. In this diagram the normal "ground state" is taken as the zero of the energy scale; and we have plotted a few of the amounts of energy that the atom will accept. (b) A similar diagram for the cesium atom.

The ground state is the state in which we normally find the atom before any excitation energy has been fed into it. Above it lie the various excited states, separated by gaps. These excited states are formed when enough energy is supplied by the impinging electrons. Finally, when an atom is hit by an electron of sufficient energy, the atom is in fact disrupted. An electron is ejected from the atom and a positive ion left behind. Since the energy of the ejected electron can take on any value, the atom can accept any parcel of energy larger than the ionization energy. For mercury, as we have indicated in Fig. 33–3, the ionization energy turns out to be 10.4 electron volts.

The atoms of many elements have been studied by electron collision. From these experiments, we find that every kind of atom has its own set of energy levels and its own ionization energy. They identify the atom much as its chemical reactions or its optical spectra identify it.

33–3 DISSECTING ATOMIC SPECTRA: EXCITATION AND EMISSION

A completely different means of measurement fully confirms the existence of energy levels. While we are bombarding a gas with electrons, we observe the light emitted by the gas. At low bombarding energies, we see no appreciable light emitted by the gas. But when we pass the first excitation energy, light suddenly appears. Examining the light with a spectroscope, we see a sharp line indicating light of a definite wavelength, photons of a single energy.

With mercury, for example, when the energy of the bombarding electrons passes 4.9 electron volts, a spectroscope with which we are observing the gas will suddenly show the presence of ultraviolet light having a wavelength of 2537 angstroms. This strong spectral line is one of the lines by which the mercury spectrum is recognized. But this single-line mercury spectrum is most unusual. From a discharge tube filled with mercury vapor we usually see many characteristic mercury spectral lines, many lines of definite wavelength which identify the element mercury. We have produced only one part of the normal mercury spectrum.

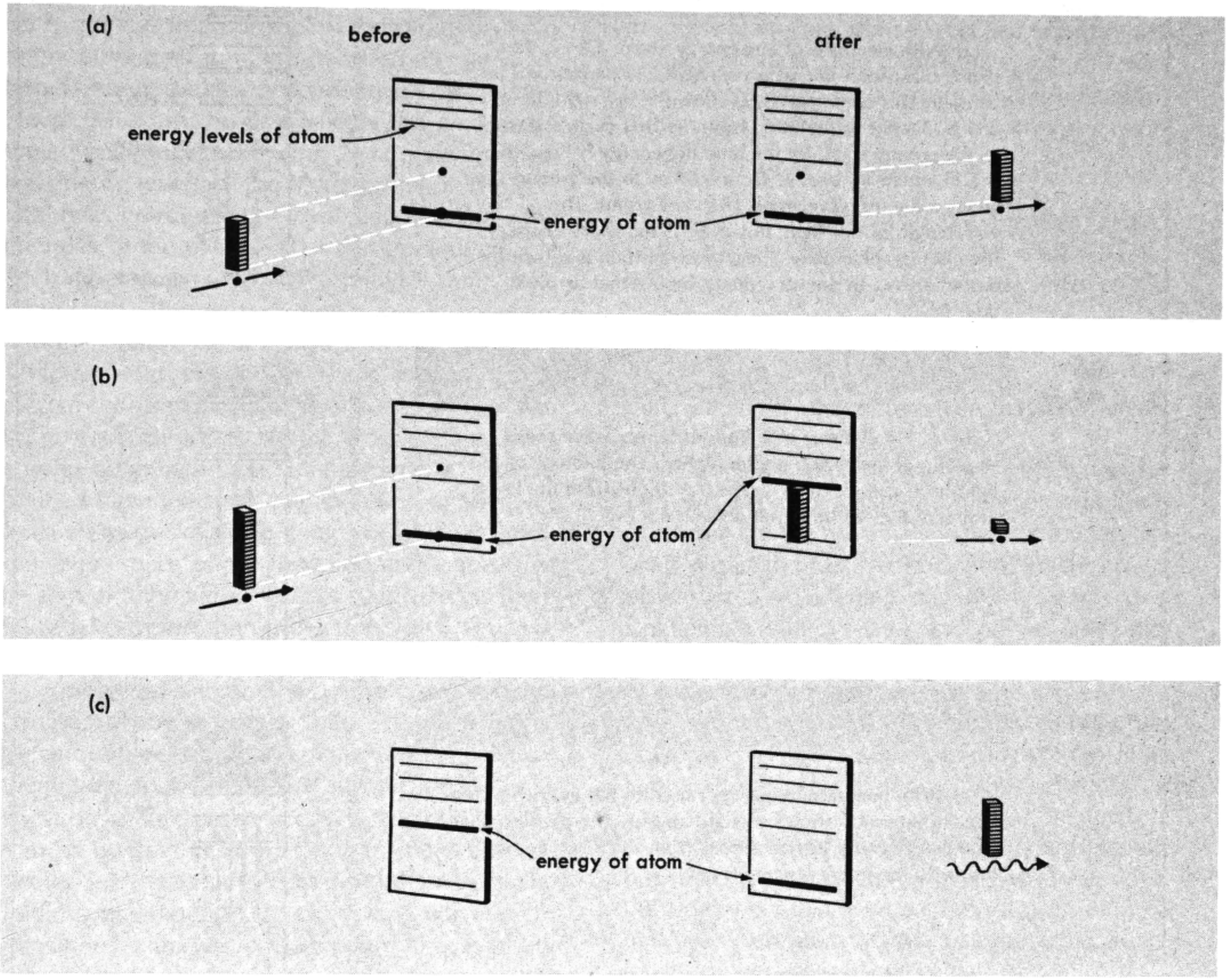
The experiments we described in the last section tell us that we have succeeded in feeding energy bundles of 4.9 electron volts to some of the mercury atoms in the gas. Only when such energy bundles have been transferred do we see any light from the mercury. Apparently the excitation energy which we feed in from the kinetic energy of the electrons is radiated as photons by the individual atoms as they return to the ground state. (Remember that the ground state is the lowest energy state, the state in which the mercury atom is normally found before the 4.9 electron volts of energy are transferred to it by a colliding electron.) If all the energy taken in by an individual atom is radiated in one lump, the photons should have 4.9 electron volts energy, and the wavelength of the corresponding light should be

$$\lambda = \frac{12{,}400 \text{ ev-Å}}{E} = \frac{12{,}400 \text{ ev-Å}}{4.9 \text{ ev}} = 2530 \text{ Å}$$

If a mercury atom could radiate its extra energy in more than one photon, light of longer wavelengths would be visible. The experiment with the spectroscope therefore shows that the atoms not only take a definite amount of energy from the electrons but that they can only radiate exactly this amount of energy in returning to the ground state. Both in accepting energy from electron bombardment and in emitting energy as light, the atoms change in one jump between two energy states a little less than 4.9 electron volts apart (Fig. 33–4). This is a magnificent confirmation of the idea that atoms can make only definite shifts of internal energy.

Let us continue to experiment with the spectroscope, increasing the energy with which the electrons bombard the mercury atoms and observing the light they emit. As we increase the bombarding energy, the rest of the familiar lines of the mercury spectrum appear. They enter in groups as the bombarding energy rises successively above the thresholds for excitation of higher and higher energy states. This procedure therefore allows us to dissect the mercury spectrum and to see how it arises.

Before an atom is hit by an electron, the atom is in its lowest internal energy state, the ground state. When an atom is hit by an electron of less than the energy difference between the ground state and its first excitation state, nothing happens. A higher-energy electron, however, may transfer energy into the atom. If there is only sufficient energy to reach the first excited atomic state, just the first excitation energy will be taken from an electron, and the atom will then radiate a photon which helps to make the 2537-angstrom line which we first saw. Many atoms contribute photons, and each photon has the same energy because each of the individual atoms falls from the first excitation state back to the ground state. When the electrons can supply still more energy, individual atoms can be excited to still higher internal energy states. Then as an atom returns to the ground state, it can radiate a photon of greater energy or possibly several photons with energies equal to the energy differences between



33–4 The changes of energy when an atom is bombarded by electrons.

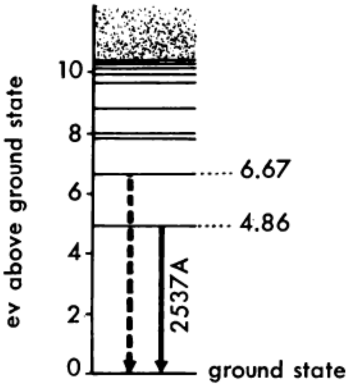
(a) An elastic collision between a bombarding electron and an atom. The arrow shows the path of the incoming electron. The "load" above the electron represents its kinetic energy. The atom and its energy levels are represented by the rectangular "billboards." In this case the electron does not have enough energy to raise the atom to the next higher energy level so it scatters off elastically, keeping its kinetic energy.

(b) Here the electron has more energy. The atom absorbs just the right amount of energy to reach its second level, and the electron goes on with only the surplus kinetic energy. The collision is inelastic.

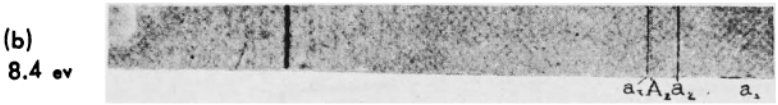
(c) The atom is at its second energy level, an "excited state." In returning to the ground state, the atom emits a photon with energy equal to the difference in energy between the two states.

The Dissection of the Spectrum of Mercury

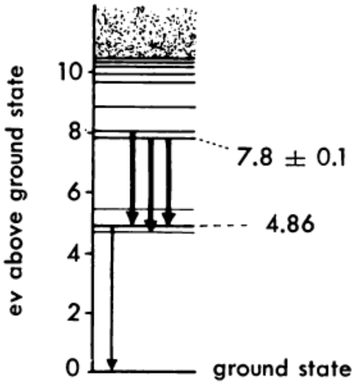
33–5 The photographs show the spectral lines that appear when mercury vapor is bombarded with electrons. In each case the bombarding electrons have a definite energy which is recorded at the side of the photograph. (Photos from: John A. Eldridge, "The Spectrum of Mercury Below Ionization," Physical Review, Vol. 23, Series 2, 1924.)


10
8
6
4
2
0
ev above ground state
6.67
4.86
2537A
ground state

(a) With electrons of any energy above 4.9 ev, the strong ultraviolet line of wavelength 2537 angstroms is emitted. This corresponds, as shown at the right, to an energy change in the atom from the first excited state to the ground state. By the time the energy is 7 ev, there is also an energy change from 6.67 ev to the ground state, giving a line of wavelength 1849 angstroms. The wavelength of this light, however, is too short to affect the photographic plate. Here this transition is shown by a dotted arrow. In the subsequent figures that arrow is omitted.

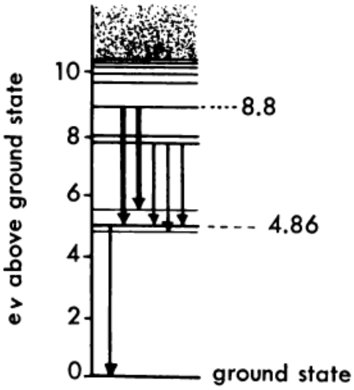

(b)
8.4 ev
a₃ A₂ a₂ a₁

(b) At 8.4 ev three new lines of longer wavelength show up in the region photographed. The photons of this light have energy given by the transitions from the levels around 7.8 ev to the levels around 4.9 ev.

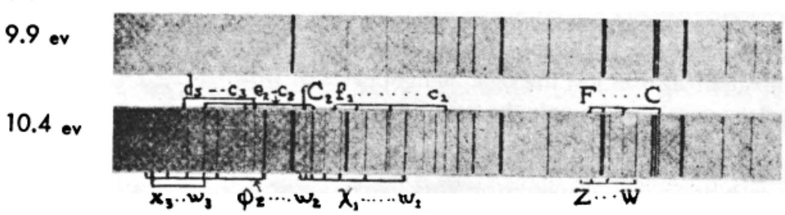

10
8
6
4
2
0
ev above ground state
7.8 ± 0.1
4.86
ground state

(c)
8.9 ev
u₃ u₂ u₁

(c) With bombarding energy raised to 8.9 ev, two more lines of intermediate wavelength appear. The photons have energies given by the transitions from the levels around 8.8 ev to the levels around 4.9 ev.


10
8
6
4
2
0
ev above ground state
8.8
4.86
ground state

(d) As the electron-accelerating voltage is further increased, more and more lines appear. At 10.4 ev the atom is ionized and the complete spectrum is produced. In this case the classifications above and below the photo help to sort out the various sets of lines.


(d)
9.9 ev
10.4 ev
d₃ -- c₃ e₂-c₂ C₂ f₁······ c₁
F···C
x₃··w₃ φ₂····w₂ χ₁·····w₁
Z···W

the several internal energy states at which the atom can exist.

In radiating photons—as in every other process that we know—energy is conserved. The energy of the emitted photon is equal to the difference between the initial internal energy and the final internal energy of the atom as it jumps from one state to another:

$$E_{\text{photon}} = E_{\text{initial}} - E_{\text{final}}$$

Each spectral line consists of the radiation from many atoms, all jumping between the same internal energy states independently of each other. Figure 33–5, in the box, shows the analysis of the mercury spectrum and its connection with the energy levels of the mercury atom. The vertical lines on the energy-level diagrams show the energies of the photons that can be emitted. For each photon emitted, some single atom must jump from a higher state down to a lower state, and the photon carries off the energy given up.

Notice that the longer wavelength lines of mercury correspond to jumps between states both of which are above the ground state. Their energy differences are smaller, so the photons emitted have lower frequency and longer wavelength. To make such long wavelength lines, however, some of the mercury atoms must first be sent into the higher energy states. To get them there takes more than 5 electron volts. Thus we can explain what otherwise would seem strange: the longer wavelength lines of mercury consist of photons of lower energy than 5 electron volts. Yet more electron energy is needed to produce them than to produce the strong line at 2537 angstroms. The reason is clear: spectral lines are formed by the transitions of atoms between particular internal energy states. And none of the initial states required for these long wavelength transitions can be reached from the mercury ground state without energy transfers well above 5 electron volts.

So far in our discussion we have assumed that the mercury atoms are normally found in the lowest energy level, or ground state. This is true at room temperature. As temperature increases, however, the collisions among mercury atoms become more violent. At sufficiently high temperature, some atomic collisions can be inelastic just as electron collisions can. They leave one or both of the colliding atoms in an excited state. So as the temperature rises, more and more atoms are moved into excited states. When enough atoms are in excited states, the gas glows. The light is emitted as the atoms rapidly return to the ground state, sometimes stepping from energy level to energy level, emitting photons as they go, sometimes jumping from a high level to a low one emitting a high-frequency photon.

High temperatures, electron collisions, and every other way of producing emission of light from gas atoms are all means of promoting atoms to excited states. The atoms then emit the photons whose energies correspond to the energy differences between the atomic energy levels.

$$E_{\text{photon}} = E_{\text{initial}} - E_{\text{final}}$$

These emitted photons make up the atom's spectrum. A given kind of atom can emit only the lines which have the characteristic wavelengths

$$\lambda = \frac{hc}{E_{\text{photon}}}$$

determined by the differences between its energy levels.

Of course, different ways of putting energy into the gas can result in different proportions of atoms in the various excited states. As we shall see in the next section, photons from light of a single wavelength can excite atoms to a single high energy state; and only a few other states may be reached as the atoms radiate their way back to the ground state. The carefully adjusted electron beam of 5 electron volts excites only the first excitation state of mercury. High temperatures act differently. The collisions of the randomly moving atoms raise the internal energies of a few atoms to high energy levels, but there are more atoms in lower states. Thus the brightness of lines can vary, but an individual line of an atom always has the same wavelength whenever it is present.

33–4 ABSORPTION SPECTRA

Emission spectra—the bright lines we see in the spectra of excited atoms—are not the only spectra which reveal the internal energy states of the atoms. We get similar information about those states when we send white light through a gas and then analyze it with a spectroscope. The white light is a mixture of all frequencies; it contains photons of all energies. Most of these photons pass through the gas and show up in the spectrum on the spectrographic plate. They are not absorbed, because the energy of the photon cannot be taken up by the atom as it makes a transition from one internal energy level to another. But some of the photons are taken up. Their absence is shown by narrow gaps in the spectrum of the light which passes through the gas. Each missing photon has been absorbed by an atom, causing a transition from one internal energy level to another. These gaps in the spectrum then tell us about the internal energy states (Fig. 33 – 6).

Here the same interpretation and the same rule of energy conservation apply; only the order of the process is different (Fig. 33 – 7). In this case the initial state of the atom is the lower internal energy state; the final state is the higher, because it must have the full energy of the original atom plus that of the absorbed photon:

$$E_{\text{final}} = E_{\text{initial}} + E_{\text{photon}}$$

In this way one can see, for example, why mercury vapor at room temperature is transparent to visible light. No photon in visible light has enough energy to promote the mercury atom from the ground state to its first excited state. The right photons come only in ultraviolet light of 2537 angstroms wavelength. As we should expect, that particular light is very strongly absorbed. The same argument applies to the other common colorless gases—oxygen, nitrogen, helium, and the rest. They can absorb light only in the ultraviolet. And there we find a series of absorption lines.

2594
2544
2512

33–6 The short-wavelength region of the absorption spectrum of sodium. Since this is a reproduction of a photographic negative, the wavelengths absorbed appear as light lines on a continuous dark background. (From: H. Kuhn, "Über-Spektren von unecht gebundenen Molekülen (Polarisationsmolekülen) K_2, Na_2, Cs_2, und Verbreiterung von Absorptionslinien," Zeitschrift für Physik, Vol. 76, 1932.)

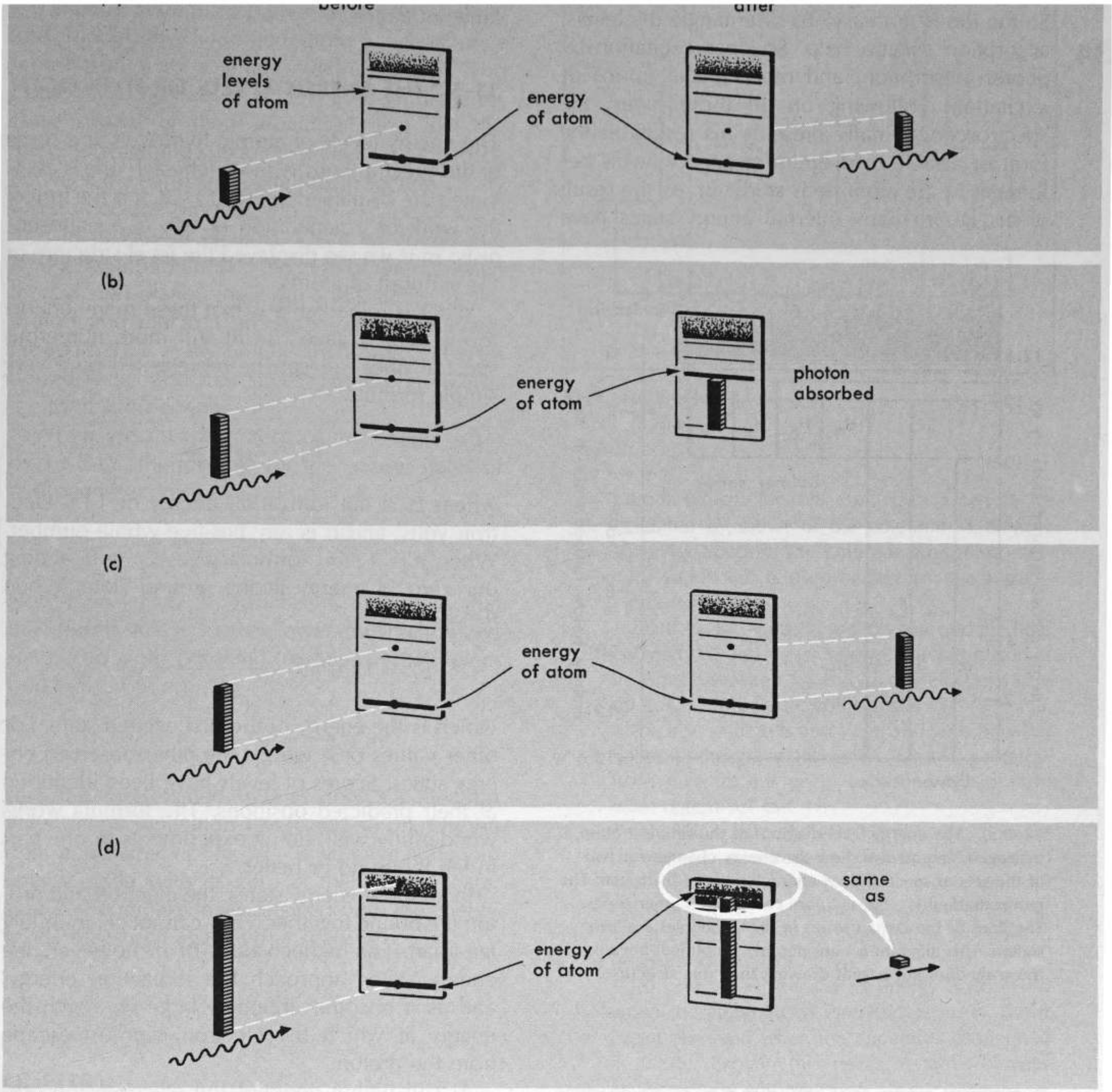
Once the frequency of the light is sufficiently high, in the deep ultraviolet or beyond, a photon can take an electron completely out of an atom. The atom is ionized, and any surplus energy is carried away by the electron as kinetic energy. This is the photoelectric effect. There is no limitation on the amount of kinetic energy the electron can be given. So there are no restrictions on the frequency of the light that ejects electrons provided it exceeds a certain minimum "threshold" value. The observed threshold of the photoelectric effect in atoms is just the minimum frequency required to break an electron loose from the atom. The evidence of spectra in this matter agrees perfectly with the evidence obtained by observing the photoelectrons.

All light, all electromagnetic radiation can excite atoms, or be emitted by them in jumps or transitions between energy levels. Consider one example far outside the range of visible light. Although we are apt to think of radio waves interacting with antennas rather than with atoms, the study of atomic spectra at radio and microwave frequencies is a modern and powerful tool for physicist and chemist. Because radio waves are long, their photons carry little energy, so they arise from levels which lie very close together in energy. All the alkali metals, lithium, potassium, rubidium, and cesium, have pairs of close-lying energy levels. The photons from transitions between these pairs of energy states are in the radio-frequency part of the spectrum. In particular, the ground state of the element cesium is really a pair of energy levels separated by about 4.14 × 10^{-5} electron volt. Precisely, the radiation from this transition between energy levels has a frequency of

$$\nu = 9.19263177 \times 10^9 \text{ cycles per second}$$

This frequency is now used as a standard of frequency and time. Clocks working directly with it are commercially available.

The spectrum has become the best clue to the energy levels of any radiating atom. But realizing this and establishing the energy levels from these clues required great ingenuity. Once the spectral wavelengths are known, they must be converted to frequencies, and the frequencies examined and fitted together to find the energy levels themselves. Each spectral line gives a difference of two energy levels, and a few levels



33–7 The change in energy when an atom is bombarded by light photons of various energies. The "load" sketched above the photon represents its energy.

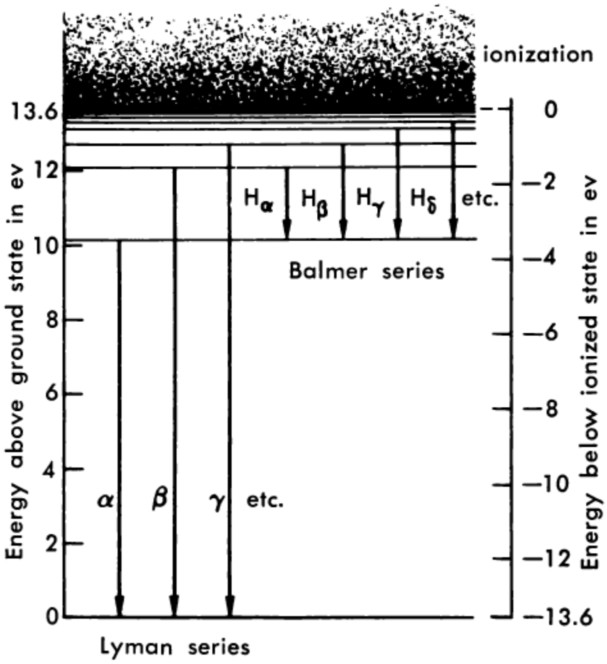
(a) The atom is in the ground state, and the photon has too little energy to change the atom to the next excited state. The photon simply passes through or scatters off elastically.

(b) The photon is one that has exactly the right energy to raise the atom to its first excited state. The atom absorbs the photon, taking on all the energy of the photon.

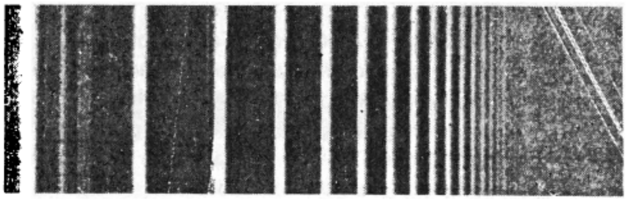
(c) The photon is one that has more than enough energy to excite the atom but not enough to reach another excited state. This photon, too, must scatter elastically, for if it were absorbed the atom could not both conserve energy and be in an allowed energy level. The only alternative is the reemission of the photon.

(d) In this case the photon is one with enough energy to ionize the atom. It raises the atom to a region filled with positive energy levels. This is another way of saying that the atom is broken up and the electron carries away the energy above the ionization threshold as kinetic energy. (Note that an atom which is already excited can be ionized by a photon of lower energy which would simply have been scattered elastically from an atom in its ground state.)

can give quite a complicated collection of lines. So the job is not easy. To disentangle the levels, absorption spectra help. So does excitation by photon absorption, and other special forms of excitation. Following up all these clues, the spectroscopist finally presents his results in the form of a table or diagram of energy levels belonging to the atom he is studying. As the result of his labor, many internal energy states have been precisely established for practically all kinds of atoms.



33–8(a) The energy level diagram of the simplest atom, hydrogen. The arrows show the energy changes for two of the sets of spectral lines characteristic of hydrogen. The jumps marked H_α, H_β, etc., are called the Balmer series. The lines of the Lyman series in the ultraviolet are also shown. This diagram is considerably simplified, because the scale does not permit drawing the "fine structure."



33–8(b) A photograph of the Balmer series, from n = 6 to as high as n = 20 or more. Their positions fit the predicted values to within one part in ten thousand or even better. (Photo from: G. Herzberg, "Uber die Spektren des Wasserstoffs, "Annalen der Physik, Vol. 84, 1927.)

33–5 THE ENERGY LEVELS OF HYDROGEN

The energy levels of atomic hydrogen and some of the spectral lines with which each one is associated are displayed in Fig. 33 – 8. It is the fruit of the work of a generation or two of spectroscopists; in it we see displayed the level structure of the simplest of atoms.

What is remarkable is that these many energy levels, which gave rise to still more numerous spectral lines, all can be represented by a very simple formula

$$E_n' = E_I - \frac{E_I}{n^2}$$

where E_I is the ionization energy of 13.6 electron volts, and n is any positive whole number. When $n = 1$, this formula gives $E_1' = 0$, setting our zero of energy at the ground state. When $n = 2$, we get

$$E_2' = 13.6 \text{ ev} - \frac{13.6 \text{ ev}}{(2)^2} = 10.2 \text{ ev}$$

which is the energy of the first excited state. For other values of n we get the other observed energy states. Scores of levels have been identified at their predicted positions. The formula works wonderfully well, fitting experiments to one part in ten thousand or better.

In all these energy states, the electron and proton are bound together. They cannot get infinitely far apart. For higher values of n, however, the energy states approach the ionization energy, and as n becomes infinitely large we reach the energy at which the electron can just escape from the proton.

At still higher energies the electron and proton are no longer bound together. The electron is just moving past the proton. At these energies, even when the electron gets infinitely far from the proton, there will be kinetic energy left over. On the energy-level diagram you will notice the essential difference between these energy states in which the electron moves freely away with any energy and the bound states (beneath the ionization energy). There are free states at all energies above ionization, while there are gaps between the individual bound energy states.

In Chapter 26 we saw that the atom resembled a planetary system. In particular, on the Rutherford model the hydrogen atom consists of a proton sun with a single electron planet. The Coulomb force between them goes with the inverse square of their separation, just like the force between the sun and a planet, and the force between the earth and a satellite. In Section 18 – 5 we studied the binding and the escape of satellites in just these systems. There we found that it was natural to choose the zero of energy when the sun and planet, the earth and satellite, or the proton and electron are standing still infinitely far apart. That energy marks the border between bound states and free states. For the atom it is the ionization energy.

When we raise the zero of the energy scale by 13.6 ev, the formula for the energy states of hydrogen becomes even simpler. It is

$$E_n = E'_n - E_i = -\frac{E_i}{n^2}$$

Now $E_1 = -13.6$ ev, which tells us directly that the ground state is 13.6 ev lower than the ionized state. The atom is bound together by 13.6 electron volts. The equality

$$E_2 = \frac{-13.6 \text{ ev}}{(2)^2} = -3.4 \text{ ev}$$

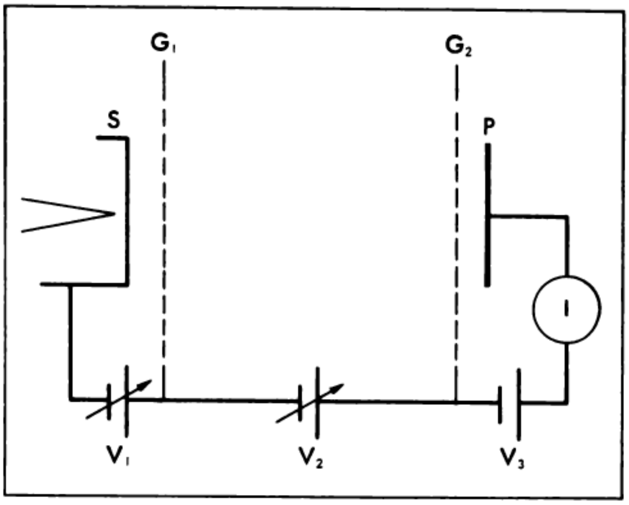
says the first excited state is only 3.4 electron volts below ionization level; and so on. On the right-hand edge of Fig. 33 – 8 we have added an energy scale with its zero at the ionized state. You can compare these answers with the energy levels measured against that scale.

QUESTIONS

1.* What does the Rutherford model of the atom predict about the ratio of the frequency of the light emitted by a hydrogen atom when the radius of the electron's orbit is halved, compared with the original frequency?

2.* An electron whose kinetic energy is 4.2 ev collides with a mercury atom. What will be the kinetic energy of the electron after the collision?

3. Figure 33 – 9 is a schematic sketch of a vacuum tube in which electrons emitted by S can be accelerated by either of the potential differences V_1 or V_2. Suppose the tube contains some mercury vapor and is used to perform the Franck-Hertz experiment.



33–9 For Problem 3.

(a) If $V_1 = 0$ volts, $V_2 = 5.0$ volts, and the field between G_1 and G_2 is uniform, in what region of the space between G_1 and G_2 can electrons give up 4.9 ev in inelastic collisions?

(b) If $V_1 = 0$ volts and $V_2 = 10.0$ volts, in what region of the space between G_1 and G_2 can electrons give up 4.9 ev in inelastic collisions? In what region 6.7 ev?

(c) If $V_1 = 10$ volts and $V_2 = 0$ volts, how far do the electrons have to go through the gas before they can lose 4.9 electron volts in inelastic collisions? How far before they can lose 6.7 ev? What electron energies should it be possible to measure to the right of G_2?

(d) The 6.7-ev transfer of energy is observed when the apparatus is operated as in (c), but it is not observed when the apparatus is operated as in (b). Explain this result in terms of your answers to (b) and (c).

4. (a) What is the significance of the constant B that appears in the equation for the photoelectric effect in Section 31 – 3?

(b) How do the values of B for the four metals exhibited in Fig. 31 – 7 compare with their ionization energies? (Ionization energies of Li, Mg, and Cu are 5.36 ev, 7.61 ev, and 7.68 ev, respectively.)

(c) What conditions seem to influence the amount of work required to remove an electron from a metal?

5. In the experiment of Franck and Hertz, schematically represented in Fig. 33 – 2, electrons are used to bombard the mercury atoms. What advantages do the electrons possess for this purpose over protons?

6.* Mercury vapor is bombarded by electrons whose kinetic energy is 6.67 ev. What are the possible energies of the emitted photons?

7.* Suppose a certain atom emits a photon with a wavelength of 6838 Å. How much energy does the atom lose?

8. Figure 33 – 10 shows a mechanical model of energy levels with platforms at four different levels above the ground. A number of uniform putty balls are located on top of *A*. When balls are pushed over the side from *A* and land somewhere below, their temperature will rise in direct proportion to their loss in potential energy. Balls which land on *B*, *C*, and *D* are again pushed until they reach the ground. How many different changes in temperature are possible? Draw the "energy levels" and mark the possible jumps.

9. Suppose you are going to bombard a vapor of cesium atoms with 4.00-electron-volt electrons. Use Fig. 33 – 3 to predict some of the things that you would expect to happen.

10. If ordinary table salt is introduced into a gas flame or a soft glass rod is heated in it, the flame is colored with the characteristic light from sodium atoms, whose wavelength is 5890 angstroms. This light is emitted by electrons returning to the ground state from the first excited energy level.

    (a) What is the energy of this level above the ground state, in electron volts?

    (b) What would be the temperature of a gas whose molecules had an average kinetic energy of this amount?

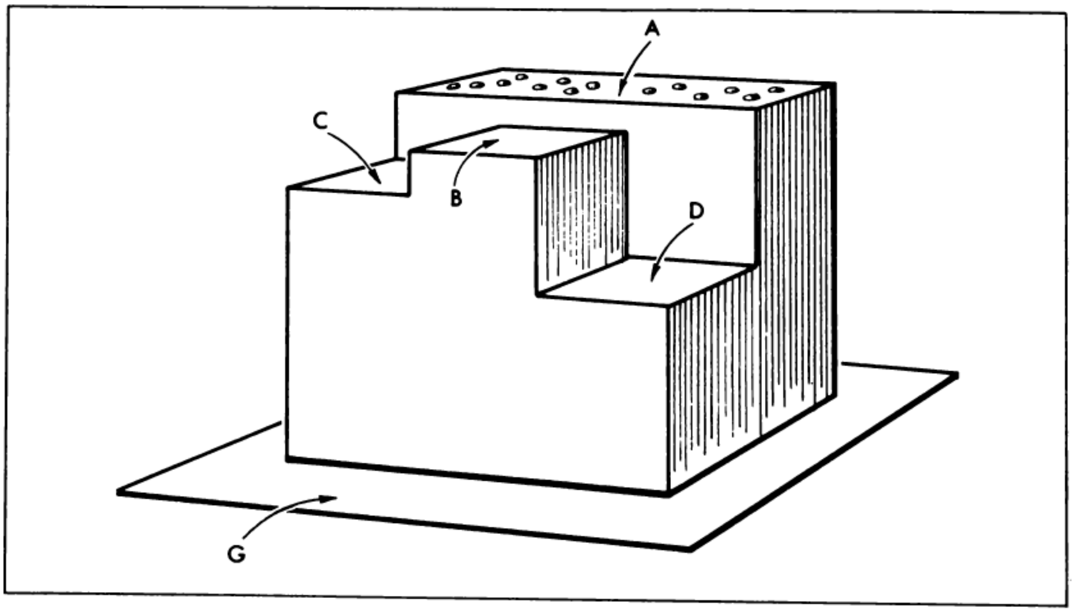
    (c) The average temperature of the molecules in a gas flame is about 2100°K. How is it possible for a gas flame to raise sodium atoms to the first excited energy level?

11. If we heat helium gas to a high enough temperature, the average kinetic energy of an atom in its thermal motion will be so high that an inelastic collision between two helium atoms can easily excite one of them to its first excited state at 19.8 electron volts above the ground state. Estimate this temperature, obtaining any data you need from Chapter 20.

12.* Compare the amount of energy absorbed when mercury vapor is bombarded with a 6-ev electron and when it is bombarded by a 6-ev light photon.

13. White light, having passed through sodium vapor, is examined with a spectroscope, and the dark lines characteristic of the absorption spectrum of sodium are observed.

    (a) What happens eventually to the energy absorbed by the sodium atoms?



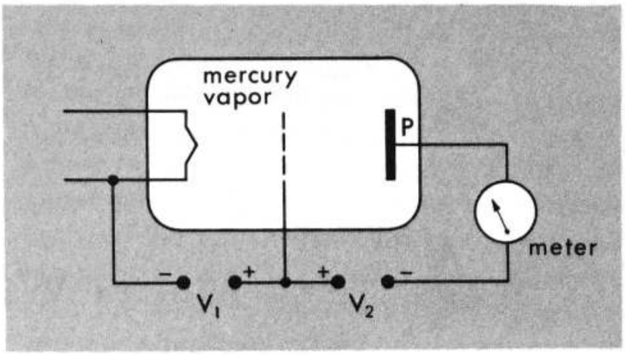
33–10 For Problem 8.

(b) Why, then, are the absorption lines dark?

14. If you point a good spectroscope at an arc lamp, you will see a continuous spectrum. If you point it out the window on a bright day, you will again see a continuous spectrum; but this time, when the slit is properly adjusted, you will see a number of dark lines crossing the spectrum parallel to the slit. (If you have access to a good spectroscope and have never done this, it is worth doing.) What is the origin of these lines? What can be deduced from a study of them?

15. Figure 33 – 11 is a schematic diagram of another vacuum tube in which electrons can be accelerated by V_1 and the energies of the electrons measured by adjusting V_2 until no current is shown on the meter. Suppose this tube contains mercury vapor.

As in the Franck-Hertz experiment, the energy of electrons arriving at P shows an abrupt decrease when V_1 is 4.9 volts. They also show an abrupt decrease when V_1 is 9.8 volts. Does this mean that mercury atoms have an energy level at 9.8 volts? Explain.



33–11 For Problem 15.

16.* If the energy of an ionized mercury atom is designated as the zero of the energy scale, what is the ground state energy of mercury?

17.* Of the seven "jumps" marked with arrows in the energy-level diagram in Fig. 33 – 8(a), which corresponds to the spectral line with the shortest wavelength?

18.* Calculate E_{10} for hydrogen when the zero of the energy scale is at (a) the ground level; (b) the ionization level.

19.* (a) With electrons of what energy would you bombard hydrogen atoms to excite them to the first excitation state shown in Fig. 33 – 8(a)? This state is 10.2 electron volts above the ground state.

(b) If you send in electrons of 12.1 electron volts kinetic energy (just enough to excite the second excited state), what energies can they lose in inelastic collisions with the hydrogen atoms?

20. In Fig. 33 – 8(b) and (c), the lines on the right side of the photo approach a limit. What is the wavelength of this limit?

21. The visible spectrum extends from 4000 to 7500 angstroms. Are all lines of the Balmer series in the spectrum of hydrogen visible?

22. (a) What energy electrons would you have to use in a Franck-Hertz experiment in order to "see" the Lyman α line? (See Fig. 33 – 8.)

(b) In order to "see" the H_α line?

(c) What other spectral lines would you expect to "see" if you can see the H_α line?

23. (a) What wavelength of light is necessary to excite hydrogen atoms in their ground state so that they emit the H_α line (Fig. 33 – 8)?

(b) What is the longest wavelength that can be absorbed by the hydrogen atoms in making a transition from the ground state (Fig. 33 – 8)?

(c) Compare this with the H_α wavelength. (You need only one really significant figure.)

34 Matter Waves

THE IDENTITY of the atoms, the stability of their internal motion, and the discrete spectrum of the light which they emit and absorb present severe difficulties for the atomic model, which cannot be resolved by a simple modification. The understanding of atoms requires a fundamental change in our approach to the subject.

To see in which direction this change has to be made, we follow the work of Louis de Broglie, who was intrigued by the fact that although a photon behaves like a particle when it collides with other particles, *where* it is likely to go is described by a wave. The energy of a photon is, from Section 31 – 3, given by

$$E = h\nu$$

Hence, the wavelength is given by

$$\lambda = \frac{c}{\nu} = \frac{hc}{E}$$

In Section 32 – 4 we have seen that a photon of energy E has a momentum $p = \frac{E}{c}$. Substituting this expression in the relation for λ we find a relation between the wavelength and momentum of a photon

$$\lambda = \frac{h}{p}$$

De Broglie asked the following question: Is it possible that this relation holds for all particles and not just for photons? The best way to answer this question is to perform an experiment with beams of various particles to see if they give rise to an interference pattern, since interference is the most characteristic property of waves. We might try, for example, to pass a stream of particles through a double slit or reflect them from some sort of grating. To see a sizable effect, the distance between the two slits (or the distance between the lines of the grating) must be of the order of magnitude of the wavelength of waves associated with the particles. Table 1 shows the magnitudes of the wavelengths associated with various moving particles, assuming the wavelength is given by the relation $\lambda = \frac{h}{p}$. As you can see, even if there is a wave associated with the motion of the small latex sphere, we cannot make a grating of a pair of slits fine enough to allow us to detect any interference effects. The best way to answer de Broglie's question is to reflect beams of electrons or slow neutrons from a crystal, which is a natural grating with a distance between lines of atoms of the order of 1Å.

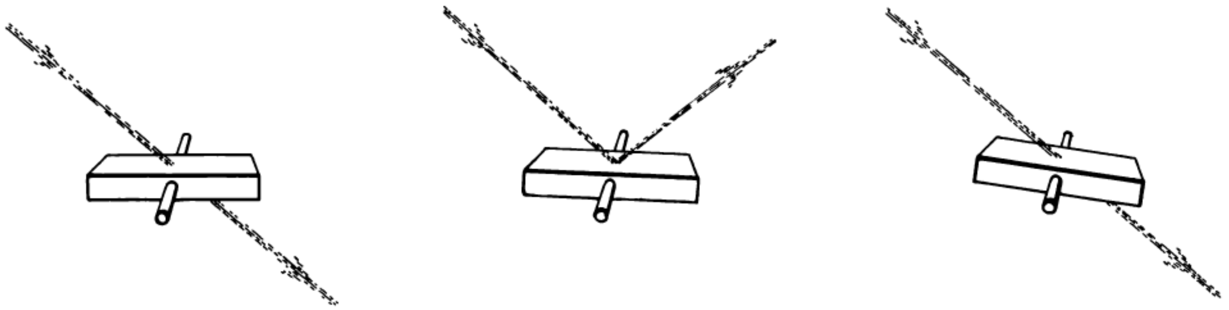
34–1 EVIDENCE FOR MATTER WAVES

If a neutron beam is directed at a salt crystal mounted so that it can be rotated about an axis, the beam will go right through it at most angles of incidence. However, if we rotate the crystal, keeping the beam direction constant, there will be an angle of incidence at which a reflected beam will suddenly appear (Fig. 34 – 1). If we rotate the crystal further, reflection will cease and the neutrons will again go through. The reflection appears when the path difference between the regularly spaced layers of the crystal gives constructive interference for the waves connected with the motion of the neutron (Fig. 34 – 2). This is analogous to the thin-film interference described in Chapter 9. Figure 34 – 3 shows an actual experimental arrangement used in an attempt to produce an interference pattern with neutrons. A beam of neutrons having a speed of 4.04×10^3 m/sec falls on the crystal from the left. When the beam is reflected, the detector responds by sending a signal to the counter for each incident neutron. The graph in Fig. 34 – 3 shows the neutron count as a function of the crystal angle as measured by the pointer on the crystal mounting. As you can see from the graph, the flash-out angle is $\theta = 10.01$ degrees. The spacing D in Fig. 34 – 2 for a salt crystal is 2.815Å. Thus $\lambda = 2d \sin \theta = 2 \times 2.815 \times 0.1738 = 0.9785$Å, which checks very closely with the predicted de Broglie wavelength of $\lambda = \dfrac{h}{p}$

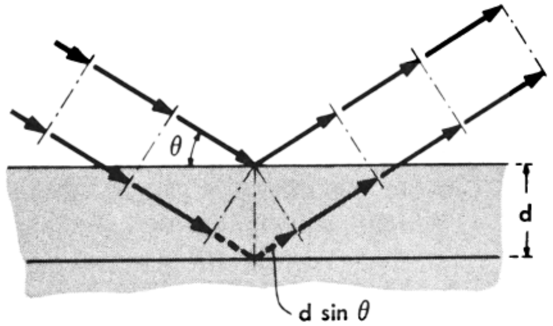
$$= \frac{6.63 \times 10^{-34} \text{ joule sec}}{1.67 \times 10^{-27} \text{ kg} \times 4.04 \times 10^3 \text{ m/sec}}$$

$= 0.981$Å.

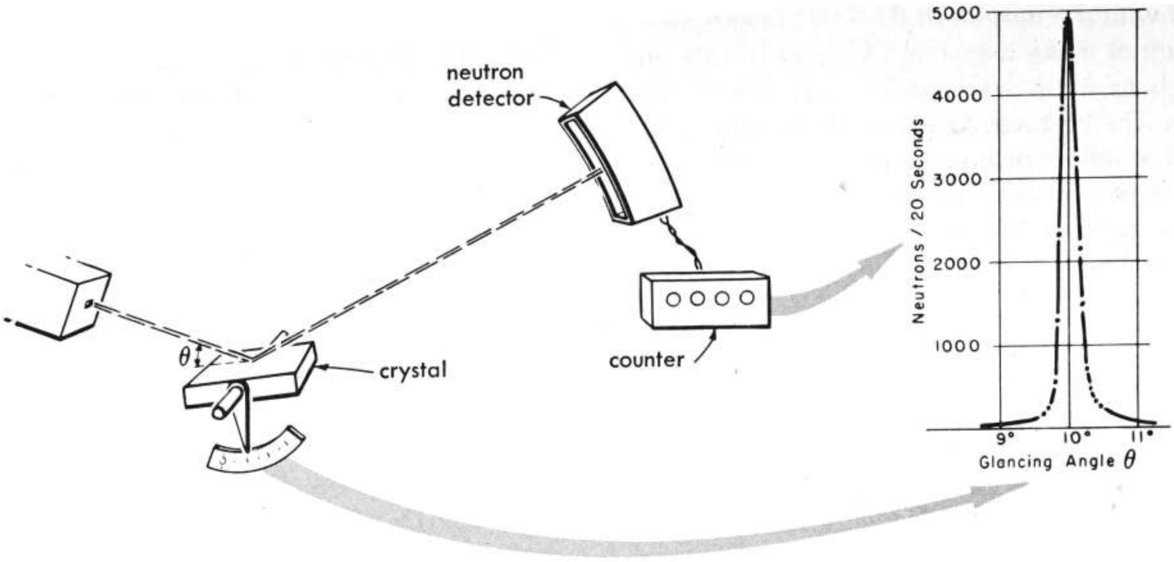
The results of such experiments with neutrons are so good that we now use neutron interference patterns to find out more about the positions of atoms in crystal gratings. Hydrogen atoms in



34–1 An interference experiment to show the wave properties of neutrons. Neutrons are reflected from a salt crystal in which the regular lines of atoms act as a diffraction grating.



34–2 The line shown dotted is the extra distance traveled by a neutron wave that reflects from the second layer of the crystal instead of the first. If the beams from each layer are to interfere to give a maximum reflection, this dotted distance must be a whole number of wavelengths. The dotted distance is also 2d sin θ, where d is the spacing between the reflecting layers of atoms. Therefore the first flash-out should occur when 2d sin $\theta = \lambda$.



34–3 The experimental arrangement for detecting the interference patterns produced by neutrons reflected from a salt crystal. Experiment by Prof. Clifford G. Shull.

crystalline compounds, for instance, whose regular spacing will not show up well in X-ray interference, form a far more effective part of the grating as seen by neutrons. Locating them in their regular repetitive array in a crystal is only one of the jobs neutrons can do because of their wave nature and their own special way of interacting with other particles of matter.

Neutrons will travel a long way in air, but experiments with atoms and electrons must be done in vacuum. Otherwise the air molecules would scatter and diffuse the beam beyond any possible use for sharp measurement. Also, when crystals are used as gratings, the particles penetrate only very slightly into the solid matter before losing their energy. Only those few particles which lose very little energy before passing out of the target material can yield a well-marked diffraction pattern. When beams of atoms are used, most of the atoms which come back from a crystal surface have in fact been somewhat disturbed internally by their collisions within the crystal, or have imparted differing fractions of their kinetic energy to the crystal. These do not make clear the wavelike behavior. In the diffracted wave there are to be found only those atoms which have behaved throughout the process as simple structureless masses, not changing in any way their internal condition. All the particles diffracted into a definite direction have the same velocity; this was directly checked to verify the point. Every moving particle, as long as it behaves as a whole and has no change internally, or in its kinetic energy, diffracts as though it has a single wavelength given by the de Broglie relation

$$\lambda = \frac{h}{p}$$

The de Broglie wavelength applies to the motion of the center of mass of the whole "particle."

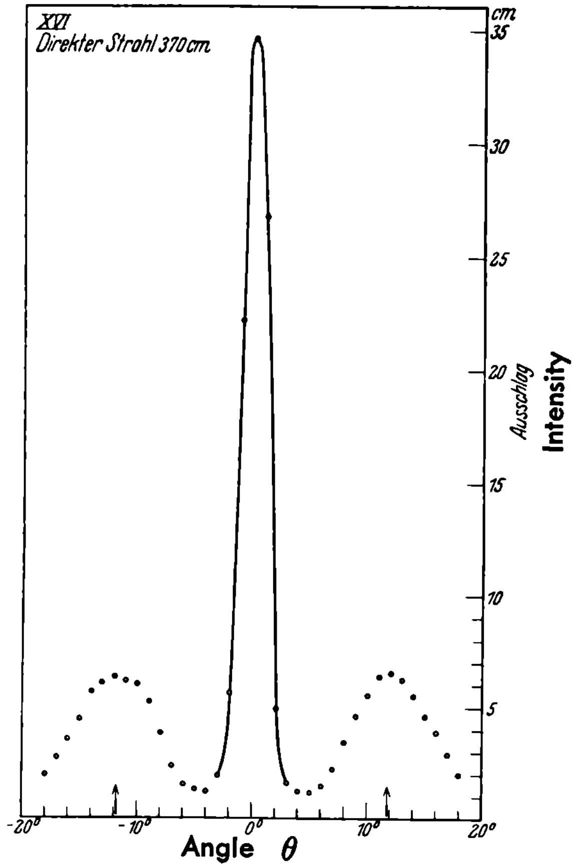
Many different kinds of interference experiments involving particles have been performed. In an early investigation of matter-wave interference helium atoms were used. Figure 34 – 4 shows the pattern they make when they reflect from the atomic grating of a lithium fluoride crystal. Here the incident beam did not have a single well-defined momentum. It was made of atoms coming out of a warm box of helium gas. Nevertheless, the interference maxima are clearly visible, humped up around the wavelength expected from the thermal energy and the corresponding momentum of the atoms. From the crystal lattice spacing of 2.85 angstrom units, and from the measured angle, about 11.8°, the de Broglie wavelength is calculated to be about 0.57Å. The de Broglie wavelength calculated from $\lambda = \frac{h}{mv}$ for helium at speeds corresponding to room temperature is also 0.57Å.

Figure 34 – 5 describes another matter wave interference experiment involving the diffraction of electrons by a "biprism."

34-4 The graph shows the result of an experiment with a beam of helium directly at a freshly cut surface of a lithium fluoride crystal. The existence of interference shows the wave nature and the angle measures the wavelength.* (From: I. Estermann and O. Stern, "Beugung von Molekularstrahlen," Zeitschrift für Physik, Volume 61, p. 107, Springer Verlag, Berlin, 1930.)

---

*The beam of helium atoms came from a sample of gas at about room temperature; not all atoms have the same speed, but they tend to cluster about some central value, which determines the angle at which the maximum is seen in the diffracted beam. From the crystal lattice spacing of 2.85 angstrom units, and from the measured angle, about 11.8°, the de Broglie wavelength is calculated to be about 0.57 angstrom unit. The de Broglie wavelength calculated from $\lambda = \frac{h}{mv}$ for helium at speeds corresponding to room temperature is also 0.57 angstrom unit.



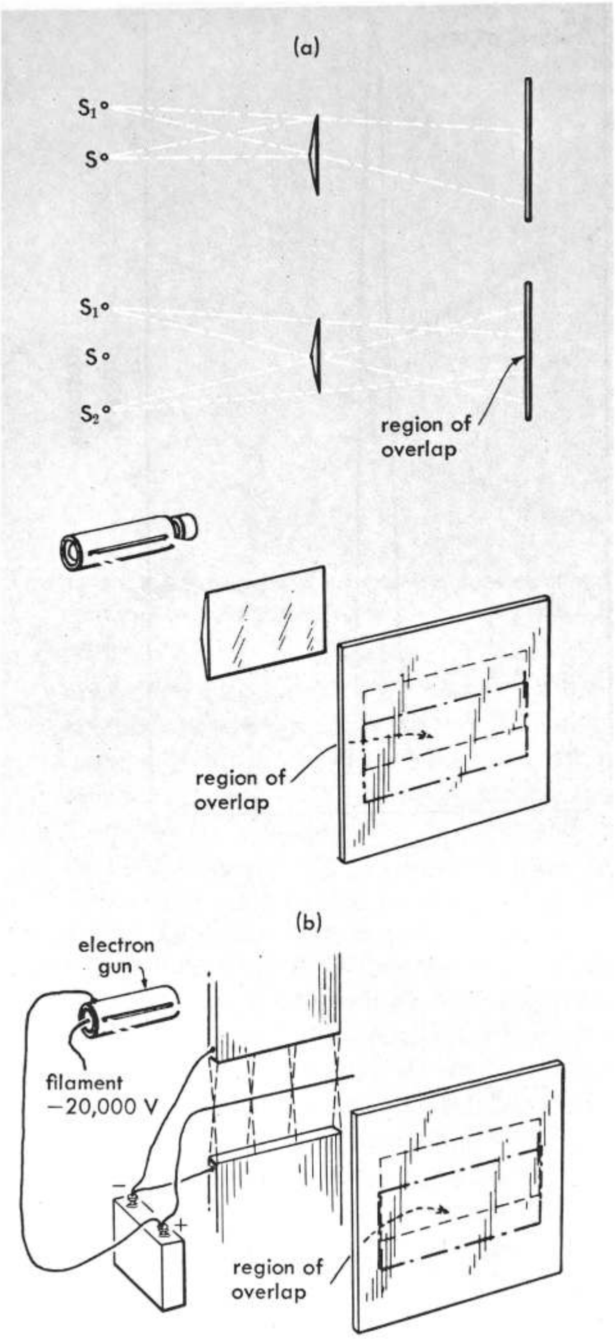
Beugung von He an Li F, 295° K,

34-2 WHEN IS THE WAVE NATURE OF MATTER IMPORTANT?

To observe effects which show us the wave nature of matter, we must have de Broglie wavelengths of the same order of size as the object we can use to diffract them. If we use moving particles whose speed or mass is large, $\lambda = h/mv$ will give us a wavelength too short to be diffracted by atomic particles. How can we lengthen the de Broglie wavelength? We can try to reduce the speed; but this is not practical. Thermal agitation itself limits the slowness of motion which can in fact be realized, but we can use smaller masses. The smallest masses are the particles of which matter is composed: the molecules, the atoms, the electrons, and the particles within the nucleus. As we have seen, their wavelengths can be long enough that we can see their diffraction patterns. Because an electron has a mass of only 10^{-30} kg, at a reasonable speed it will have a wavelength of the size of an atom. An electron moving with 100 electron volts of kinetic energy moves at a speed of 5.9×10^6 meters per second and its de Broglie wavelength is therefore about 1.3×10^{-10} meter (Table 1), which is in the range of atomic dimensions. A reasonably slow-moving electron with one electron volt of kinetic energy moves at about a tenth of the speed, and has a wavelength of about 10 atomic dimensions. Atoms are small obstacles for such

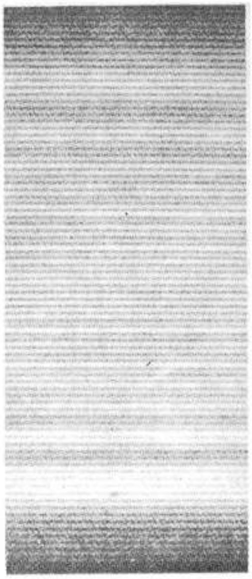
TABLE 1

PARTICLE	MASS (kg)	SPEED (m/sec)	WAVELENGTH $\lambda = \frac{h}{mv}$ (Å)
Marble	2.0×10^{-2}	1.0×10^{-2}	3.3×10^{-20}
Latex sphere	1.0×10^{-15}	3.0×10^{-4}	2.2×10^{-5}
10 Mev neutron	1.7×10^{-27}	4.4×10^{7}	0.88×10^{-4}
1 Mev proton	1.7×10^{-27}	1.4×10^{7}	2.8×10^{-4}
100 ev electron	9.1×10^{-31}	5.9×10^{6}	1.2
0.1 ev neutron	1.7×10^{-27}	4.3×10^{3}	0.90



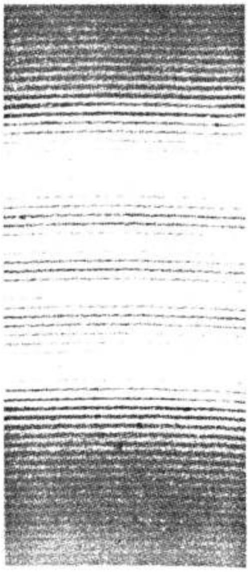
34–5 Two analogous experiments showing the interference of waves. Both show interference fringes from a biprism — a device that takes waves from a single slit as the source and produces the effect of waves from two slits. The likeness of the patterns obtained with light and with electrons is striking. The bending would happen in just the same way if electrons behaved simply as charged particles; but the peculiarity is that where the two streams of electrons overlap there are bright and dark bands just like those of light waves, though on a very small scale.

(a) With visible light, a glass biprism (two wedges of glass pointing opposite ways) bends the two halves of the beam of light from the original slit so that they overlap on a remote photographic film. The light passing through the upper half is bent down and the lower half bent up so that in the region of overlap on the photographic film light appears to come from two slits S_1 and S_2. Interference is observed in this region of overlap. At some places on the photographic plate, crests that seem to come from S_1 arrive along with troughs that seem to come from S_2, making a dark fringe. At other places crests arrive together, making a bright fringe.



(Above) The interference pattern painted by light. (From: Valasek, Introduction to Theoretical and Experimental Optics, John Wiley & Sons, New York, 1949.)

(b) The electron source is a 20,000-volt gun which provides electrons with a de Broglie wavelength of 0.086 angstrom. The electric field "biprism" is made by a fine silvered quartz thread, a few microns in diameter, suspended in a wide slit in a metal screen. The thread is kept about 6 volts positive with respect to the slit. The diagrams show the electrostatic biprism and its electric field lines. The electric field bends the streams of electrons that pass each side of the wire. They overlap on the photographic film, arriving there as if they came from two sources.

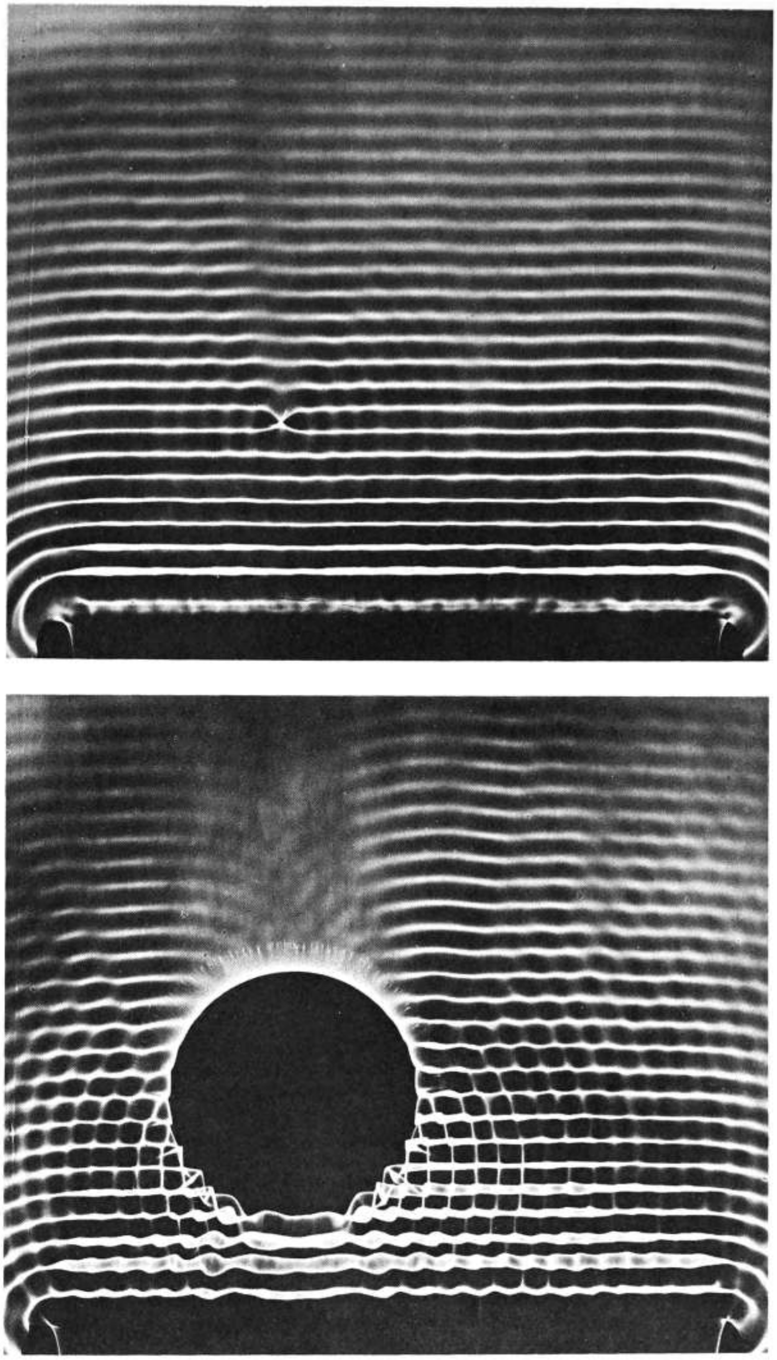


(Above) The interference pattern painted by electrons. It has been hugely magnified by an electron microscope. (From: H. Duker, Lichtstarke Interferenzen mit einem Biprisma fur Elektronenwellen," Zeitschrift fur Naturforschung, Volume 10A, 1955.)

an electron. The associated waves can easily diffract around them (Fig. 34 – 6). These examples show that wave behavior is essential whenever an electron of small kinetic energy moves in or near an atom. For these motions our ordinary experience with Newton's particle mechanics is not an adequate guide.

The situation is the same for protons and neutrons with respect to the size of the atomic nucleus. The wavelength of a proton or neutron with 100 electron volts of kinetic energy is about 3×10^{-12} meter. This wavelength is large compared to nuclear dimensions, which are 10^{-14} meter or less. Consequently, such a neutron approaching the nucleus from the outside will find it a small obstacle around which it diffracts. Neutrons with ten million electron volts have a wavelength of 10^{-14} meter, comparable to the nuclear size. Consequently, neutrons of this energy just get inside the nucleus. A neutron that stays inside a nucleus, forming a part of it with practically no chance of being found outside, must have a wavelength of this size or smaller. Consequently, we expect that the neutrons and protons in nuclei have kinetic energies of about ten million electron volts or more. Even at these energies their wave nature is important as they move about inside the nucleus.

Once again we have found that the atomic world cannot be fully described by the simple models of the older physics. The little particles cannot be followed like so many bullets as they fly about within the atom; they obey rather the laws followed by waves. The waves determine the probable location of the particles, and the motion of wave trains determines the probable motion of the particles. Where the amplitude of the wave is large, there the particle is probably to be found; where it is small, there it will be found rarely. But all we can learn is a probability. Exactly where the electrons will be found in an atom we cannot predict. Inside the atom we must be prepared to use a new version of mechanics which combines parts of Newton's particle mechanics with the effects of the wave nature of matter.



34–6 Ripple-tank photographs of waves meeting obstacles. In the upper photo a small object is located where the waves seem to bend together. Notice that this small barrier disturbs the waves very little. In the lower photo there is a large obstacle in the ripple tank. The waves are strongly reflected and there is a definite "shadow" region.

34-3 LIGHT AND MATTER

The circle has closed. Waves of light proved to describe the probability of the appearance of certain particlelike photons; the particles of matter turn out to be governed by a wavelike quantity. Planck's constant controls the magnitudes involved in both cases. We cannot follow the path of a particle through the slits of a diffraction grating any better than we can follow the path of a photon. The associated waves diffract and from these waves we can calculate probable locations of a particle and probable numbers of arrivals. But, just as in the photon picture, the particles transfer energy and momentum in every single collision in the same way as they are expected to do from the mechanics of Newton. The wave nature does not modify that at all. Therefore, whenever the wavelengths are small compared to the dimensions of slits or obstacles, small compared to the dimensions of atoms or of measuring instruments, we get results identical with the predictions of ordinary mechanics. Indeed, it is for this reason that we can use ordinary measuring instruments, and it is for this reason that ordinary measurements will only allow us to measure the same quantities which are dealt with in Newtonian mechanics.

For all objects of everyday size, or for objects like the planets, the mechanics of Newton is without any important error, but if we look with great refinement, or if we ask questions on the scale of the atom, the classical physics (as physics before 1900 is often called) is found wanting. It remains the foundation of physics, because the very instruments and apparatus with which we measure are well understood in classical terms, but it is powerless to explain the atom. Only the quantum theory can do that, with its orders of magnitude proclaimed by Planck's constant, and its more subtle picture of the atomic world. The photon governed in space and time by its electromagnetic wave of probability, the particle by its matter wave of probability—all with the de Broglie wavelength—are the keys with which we propose to enter the interior of the atom.

In all that we have said, the analogies between the behavior of light and matter have been striking. Light behaves both as waves and as particles —photons. Matter behaves both as particles and as waves. But the similarities have their limitations. There is one sharp distinction between a photon and a particle of the sort we usually think of, between a photon and a little bit of matter. Photons are not conserved in number; they are easily created and easily destroyed. To turn on the light creates photons by the swarm; they rush to the walls. Some are reflected, but in the end all die, giving their energy to the atoms they strike. Energy is conserved, for heat or chemical changes will reveal the photon energy content, but the number of "particles" is not. This behavior is sharply different from the behavior of electrons and atoms. Those particles cannot be created or destroyed (not in any commonplace way, at least). They are permanent as a photon is not. In this sense photons differ very much from little grains of matter.

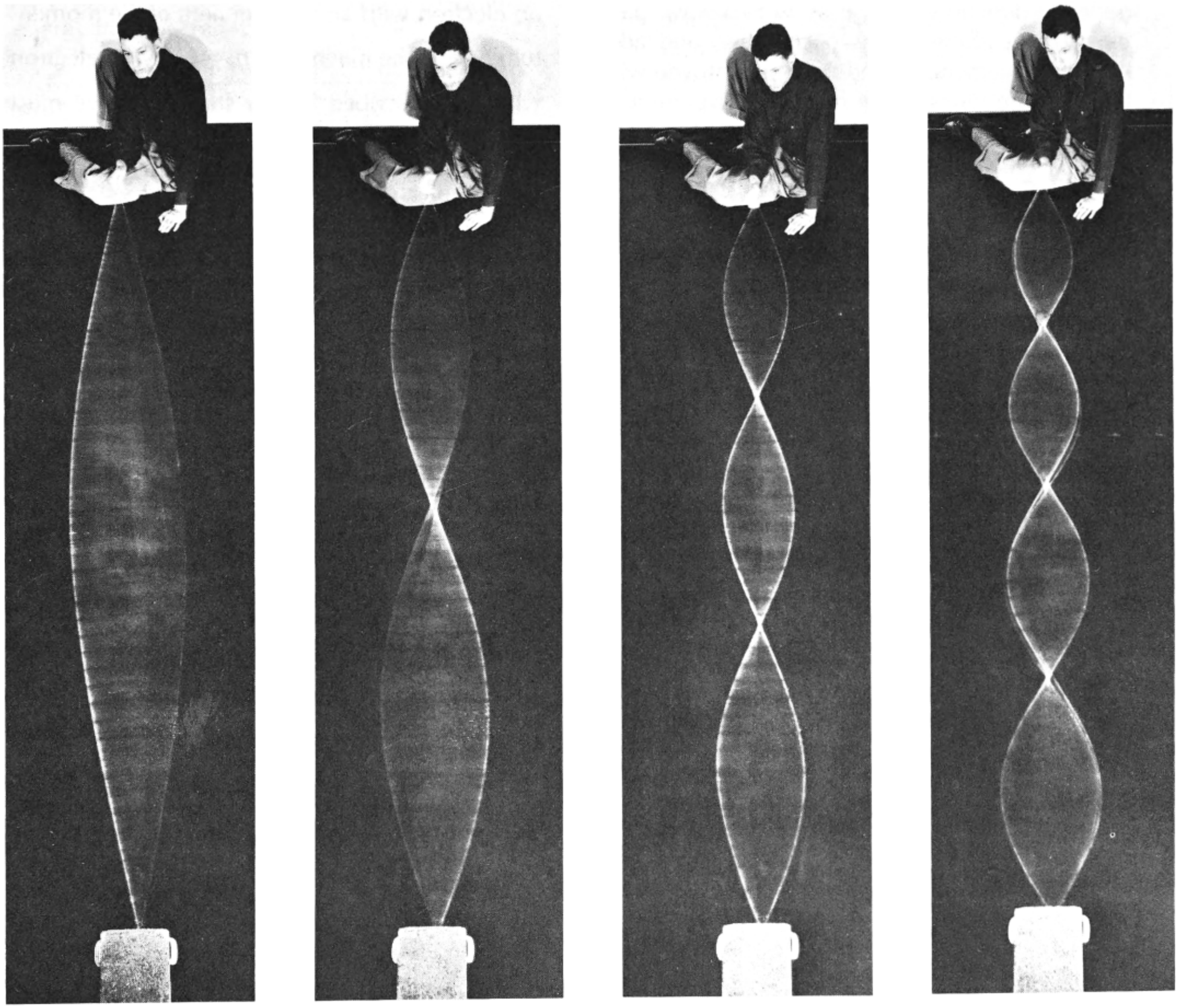
In physics today we regard the wave-particle relationship as fundamental for every one of the objects we deal with on the atomic scale; the main distinction between light and matter is between the permanence of particles and easy creation or absorption of photons.

34-4 STANDING WAVES

How can we understand the presence of definite internal energy levels of atoms and the simple numerical relations they sometimes have? We hope to do this by using the ideas of this chapter, namely the idea of matter waves, whose intensity at any place indicates the probability of finding a particle there. How does that idea explain separate, well-defined energy levels?

We know from Franck-Hertz experiments and from the definite line spectra of elements that there are definite separate energy states in atoms. We should not find that behavior if the matter wave for a particular energy level wobbled around and did not remain the same for some time. Also, there must be a different matter wave corresponding to each separate energy level, because the energy levels are different. We must therefore look for several distinct wave patterns, each one different from the others and each unchanging. The only waves which show the same patterns all the time and do not travel away are standing waves. These are the stationary waves that you find on vibrating strings or in tubs of water. They do not move about. Instead, they stay in place and keep their shape while the displacements periodically grow and collapse without moving along. Look at Fig. 34 – 7, for example, at the second picture from the left. At the two ends of the tube and in the middle there are nodes—points which do not move. In between each pair of nodes are regions of maximum displacement. These regions always stay at the same place along the tube, but as *time* goes on the displacement ranges from maximum in one direction through zero to maximum in the other direction. Everywhere, the displacement oscillates back and forth in the same way, but nothing appears to move along the tube.

There is no substitute for making standing waves yourself. Tie a rubber tube between two firm supports. Wiggle the tube up and down with your hand near one support. By choosing the right frequency for the motion of your hand, you can set up a standing wave pattern in which the tube oscillates up and down but the pattern does not travel along. It may be a pattern with one loop between the end supports, or two loops, three loops, four loops, and so on (Fig. 34 – 7).



34–7 Standing waves. As one end of the tube is moved from side to side with increasing frequency, patterns with more and more loops are formed. Note, however, that there are only certain definite frequencies that will produce fixed patterns.

This is the kind of wave motion we are going to picture for a stable state of an atom.

With the ends of a rubber tube held fixed, only the patterns with a whole number of loops (and corresponding definite frequencies) will stand on the tube. If you try to drive the tube with any other frequency, you will find that the motion of your hand undoes some of the tube's existing motion rather than being in step and increasing the motion. Unless you use the right frequency, your efforts produce very little effect. You fail to build up a standing wave.

Although standing waves just oscillate up and down, we can view them as the superposition of two equal traveling wave trains moving in opposite directions. If you draw two wave patterns, move them across each other, and add their displacements to find the resultant, you will find that the loops of the standing wave are just $\frac{1}{2}\lambda$ long, where λ is the wavelength of either wave train. We did something like this in Section 6 – 3.

On this view the standing waves between two supports are just like standing waves you have already seen in front of a reflecting wall in a ripple tank. There, you can drive the wave toward the wall with any frequency; and the reflected waves returning from the wall combine with your forward waves to make standing waves. However, for a given position of driver and wall you must choose the "right" frequency or you will find you have to give the driver a large motion to get a big effect. With the right frequency, a very small motion of the driver builds up a large effect.

As you can see from Fig. 34 – 7, if we tie one end of a rubber tube to a rigid support and wiggle the other end very slightly with the "right" frequency, a standing wave is established. We may imagine that two rigid reflecting walls are suddenly installed at nodes in the standing wave pattern. Because we install the walls at places of no motion, they would trap a section of standing wave between them. In any such section we would see a pattern of oscillations, a standing wave, instead of a moving wave pattern. To trap a standing wave pattern the supports must be one half-wavelength apart, or two half-wavelengths, or three, and so on. We shall now use standing waves in discussing a model for the atom.

34–5 A PARTICLE IN A "BOX"

We can also imagine a standing matter wave along the x axis confined between two points beyond which its amplitude is always zero. Suppose this matter wave is associated with an electron; then the behavior of this electron would be characterized as follows:

(a) The x coordinate of the electron is always between the points A and B, since we never find the electron outside that range.

(b) We can look at the standing wave as the superposition of two traveling waves of the same wavelength and moving in opposite directions. Both of these traveling waves together describe an electron with an x component of the momentum having the magnitude $p_x = \frac{h}{\lambda_x}$. The electron which is described by the standing wave must have a component of the momentum which is both $+p_x$ and $-p_x$ at the same time. Within the context of Newtonian physics such a statement makes no sense, but then, the wave nature of the motion of a particle is important precisely when the Newtonian picture fails. Thus, in the quantum world a particle can have two equal and opposite momenta at the same time; however, since its energy is $\frac{p^2}{2m}$, its energy is not zero.

The energy of a particle described by a standing wave equals the energy of the same particle described by a traveling wave of the same wavelength. But, unlike the traveling wave, the standing wave which vanishes outside a given range can have only certain discrete wavelengths and hence only discrete energies.

Consider now a particle of mass m confined to a limited volume. To get a first idea of the effects of this confinement on the energy of the particle, we can ignore the exact nature of the force field which binds the particle between two points A and B, and assume that there is no force at all acting on the particle inside the volume but an extremely strong force acting on it at the boundaries A and B of the volume and keeping it inside. Furthermore, we shall take the volume to be a cube of side length d. This highly idealized picture is usually called a particle in a "box."

Taking the potential energy reference level as zero inside the box, the total energy of the electron is all kinetic energy:

$$E = \frac{p_x^2 + p_y^2 + p_z^2}{2m}$$

The electron cannot escape from the box, so its matter wave must be a three-dimensional standing wave. Such a wave can have different wavelengths along the three axes of the box. For a cubical box of edge length d, they must fulfill the conditions

$$\frac{n_x\lambda_x}{2} = d, \ \frac{n_y\lambda_y}{2} = d, \text{ and } \frac{n_z\lambda_z}{2} = d$$

The three wavelengths λ_x, λ_y, and λ_z then determine the three components of the momentum, $p_x = h/\lambda_x$; $p_y = h/\lambda_y$; $p_z = h/\lambda_z$, and the kinetic energy then becomes

$$E = \frac{p_x^2 + p_y^2 + p_z^2}{2m} = \frac{(n_x^2 + n_y^2 + n_z^2)h^2}{8md^2}$$

The numbers n_x, n_y, and n_z, called quantum numbers, are completely independent of one another and can each take the values 1, 2, 3, etc.

The energy levels which arise from such an analysis clearly depend on the mass of the particle and the size of the box. The quantity $(h^2/8md^2)$ is a measure of the order of magnitude of the energy scale of the system and the approximate energy difference between adjacent energy levels. For an electron ($m = 10^{-30}$ kg) in a box the size of an atom ($d = 10^{-10}$m), this factor is numerically equal to about 30 electron volts, about what you would expect for an atomic-sized system.

For the ground state, the quantum numbers have their smallest values ($n_x = n_y = n_z = 1$) and the system has a minimum energy

$$E_0 = 3h^2/8md^2$$

Now when we think back to the hydrogen atom, we can see that we are following a promising road. For just such reasons, the hydrogen atom must have a ground state—although Newtonian mechanics would allow it indefinitely lower energies. Thus an explanation of the stability of atoms is already in sight. Furthermore, there are many matter waves of higher energy that can stand in the box. There are the standing waves with 2 loops, 3 loops . . . n loops, and so on. These stationary wave states, with their definite values of energy, provide a model for the excited states of the atom. We shall discuss detailed applications in the next chapter.

QUESTIONS

1.* What is the wavelength of an electron whose speed is 1.0×10^7 m/sec?

2.* As the angle θ in Fig. 34 – 2 is increased, the beam will again reflect very strongly out of the crystal. What must $2d \sin \theta$ be equal to at this angle?

3. The layers of atoms in crystals are known (from density and Avogadro's number) to be a few angstroms apart, say 3×10^{-10} m. X rays from ordinary commercial X-ray tubes are found to be diffracted at quite large angles, such as 10°, by such crystals acting as interference gratings. Make a rough estimate of the energy of photons in such X rays.

4. Electrons passing through a double slit (fantastic but possible, as discussed in the next question) make an interference pattern on a remote screen. If the gun voltage used to provide the original stream of electrons with kinetic energy is changed from 50 volts to 5000, what does this do to the spacing of the fringes?

5. A narrow stream of 100-ev electrons is fired at two parallel slits very close together. The distance between the slits is estimated to be 10 angstrom units. The electrons passing through the slits reach a screen 3 meters away, and form a pattern of interference fringes.

   [It seems almost impossible to make such a pair of slits in any real sheet of matter, made of atoms which are themselves 1 angstrom or more in diameter, and even more impossible to measure their separation, since light has a wavelength of thousands of angstroms. Yet the equivalent of this pair of slits can be made—has been made and used. It is the "biprism" of Fig. 34 – 5(b).]

   (a) Estimate the distance between one bright fringe and the next.

   (b) The bright fringes would appear as bright marks on a fluorescent screen sufficiently sensitive to glow when bombarded with 100-ev electrons (or as dark marks on a photographic film, or as large pointer readings on some meter attached to an electron collector). What is the essential difference between the bright-fringe electrons and those that reach the screen in a dark fringe? Are they bigger? More massive? Of better quality? Of greater charge, or what?

6. In a certain type of vacuum tube, electrons escaping from the hot cathode pass through a grid maintained at a potential of −10 volts at a distance of 0.5 mm from the cathode, then hit a plate maintained at a potential of +110 volts at a distance of 2.0 mm from the cathode.

   Must the wave behavior of the electrons be taken into account in describing their passage from the cathode to the plate? Explain.

7. (a) What is the wavelength of X rays whose photons each carry 40,000 electron volts of energy?

   (b) At about what energy do electrons have a de Broglie wavelength equal to that of 40,000-volt X rays? (Give your answer in electron volts.)

   (c) What energy baseballs?

   (d) What is the wavelength of a baseball moving at 10 meters per second?

8. The sizes of atomic nuclei are hard to measure with alpha particles, because we have to look for small differences from Coulomb scattering. Neutrons are often used.

   (a) What are the de Broglie wavelengths of neutrons with kinetic energy 10^4 electron volts? 10^6 electron volts? 10^8 electron volts?

   (b) Compare these wavelengths with the diameter of a gold nucleus (15×10^{-15} m).

   (c) Look back at Fig. 34 – 6 to see how waves behave when they meet an obstacle. Suppose we fire a stream of neutrons at a small target of gold. For which of the neutron waves in (a) do you expect to find

   (i) most of the neutrons pass by the nuclei undisturbed while some are absorbed and some bounce off, all like balls in Newtonian mechanics?

   (ii) most of the neutrons pass the nucleus apparently undisturbed and the rest come out of the gold target equally in all directions, regardless of the shape of the nucleus?

   (d) With neutrons of which energy would you prefer to try measuring the nuclear size?

9. The relation between photon energy in electron volts and the wavelength of light in angstroms is shown in Section 31 – 3.

   (a) Work out a similar relation between the kinetic energy of a particle and its de Broglie wavelength.

   (b) If the particle is an electron and its kinetic energy is generated by accelerating it through a potential difference, relate the potential difference in volts to the de Broglie wavelength of the electron in angstroms.

   (c) What is the energy of electrons whose wavelength is the same as that of yellow light?

10. Borrow a cymbal from the school band (or use your own). Hold the cymbal with the handle down and strike the rim with your finger or a narrow rubber hammer. With a little practice, you should be able to form standing waves along the rim. Describe them.

11. Waves travel at 2 meters/sec in a spring fastened firmly at both ends to rigid walls 8 m apart.

   (a) What frequencies of vibration could be found in standing waves formed on this spring?

   (b) Could you produce standing waves corresponding to all these frequencies by shaking the spring at its midpoint?

12.* When a length d is chosen for the "box" discussed in Section 34 – 6, does this restrict the particle in the "box" to only one fixed energy?

13.* What is the order of magnitude of the kinetic energy of an electron (mass 10^{-30} kg) bouncing back and forth in a "box" of length 1 Å?

14. (a) Estimate the kinetic energy of a neutron in a nucleus of radius

$$R = 6 \times 10^{-15} \text{ m}$$

   (This is a big nucleus.) Assuming the nucleus is a box of length R will be good enough.

   (b) If the nucleus is smaller, should the kinetic energy be smaller?

15. Suppose that an electron is confined within a small cubic crystal about 1 micron (10^{-6} m) on a side. Would quantum effects be detectable if you observed changes in the energy of the electron? (Hint: Calculate the lowest possible energy, considering the electron as a standing wave within the crystal.)

16. Suppose the particle between fixed walls (Section 34 – 5) is charged and can, therefore, radiate. What is the frequency of its radiation as it goes from the first excited state to the ground state? Express your answer in terms of h, m, and d.

17. Suppose an electron moves so as to remain always in one plane within a square of side length d.

   (a) How would its minimum energy compare with that of an electron confined to a cubical box of edge length d?

   (b) Compare the wavelengths of the radiation emitted when each of these electrons makes a transition from its first excited state to its ground state.

18. Suppose the particle described in Section 34 – 6 were confined in a rectangular box whose dimensions were *d*, *d*, and 2*d*. How would the minimum energy of the system compare with that when it is confined to a cube of side *d*?

19. In Question 24 of Chapter 26, you calculated that the minimum possible distance of closest approach of a proton of 1.0×10^{-12} joule kinetic energy to a gold nucleus was 1.8×10^{-14} meter. In the light of Section 34 – 2, should this conclusion be modified in any way to take into account the wave properties of the proton?

35

Atoms, Molecules, and Nuclei

IN THIS CHAPTER we shall briefly discuss the main ideas of the quantum theory of the energy levels of hydrogen atoms, and then try to extend them in order to gain at least a qualitative understanding of the behavior of other atoms as well as of molecules and nuclei.

35–1 HYDROGENLIKE ATOMS: ENERGY LEVELS AND SIZE

Our first step is to look at the motion of a single electron moving in the electric field of a nucleus containing Z elementary charges. We refer to such a system as a hydrogenlike atom. From the point of view of Newtonian mechanics the hydrogenlike atom resembles the earth-moon system: at a distance r from the nucleus the force acting on the electron has a magnitude kZ/r^2.* If the electron is to move in a circle at that distance with speed v, the electric force must equal the centripetal force necessary to keep the electron in the circular orbit:

$$\frac{kZ}{r^2}=\frac{mv^2}{r}$$

Multiplying both sides of this equation by $\frac{1}{2}r$ yields

$$\frac{1}{2}\frac{kZ}{r}=\frac{1}{2}mv^2=E_K$$

You may recall that the potential energy of an electron in the field of the positively charged nucleus is

$$U_r=-\frac{kZ}{r}$$

Thus, for an electron in a circular orbit we find that

$$-\frac{1}{2}U_r=E_K$$

---

*Since we use the elementary charge as a unit of charge and the number of charges is a pure number, we do not write out kZ (elem. ch.)$^2/r^2$ to simplify the notation.

Using this relation, we can express the total energy of the electron in two forms

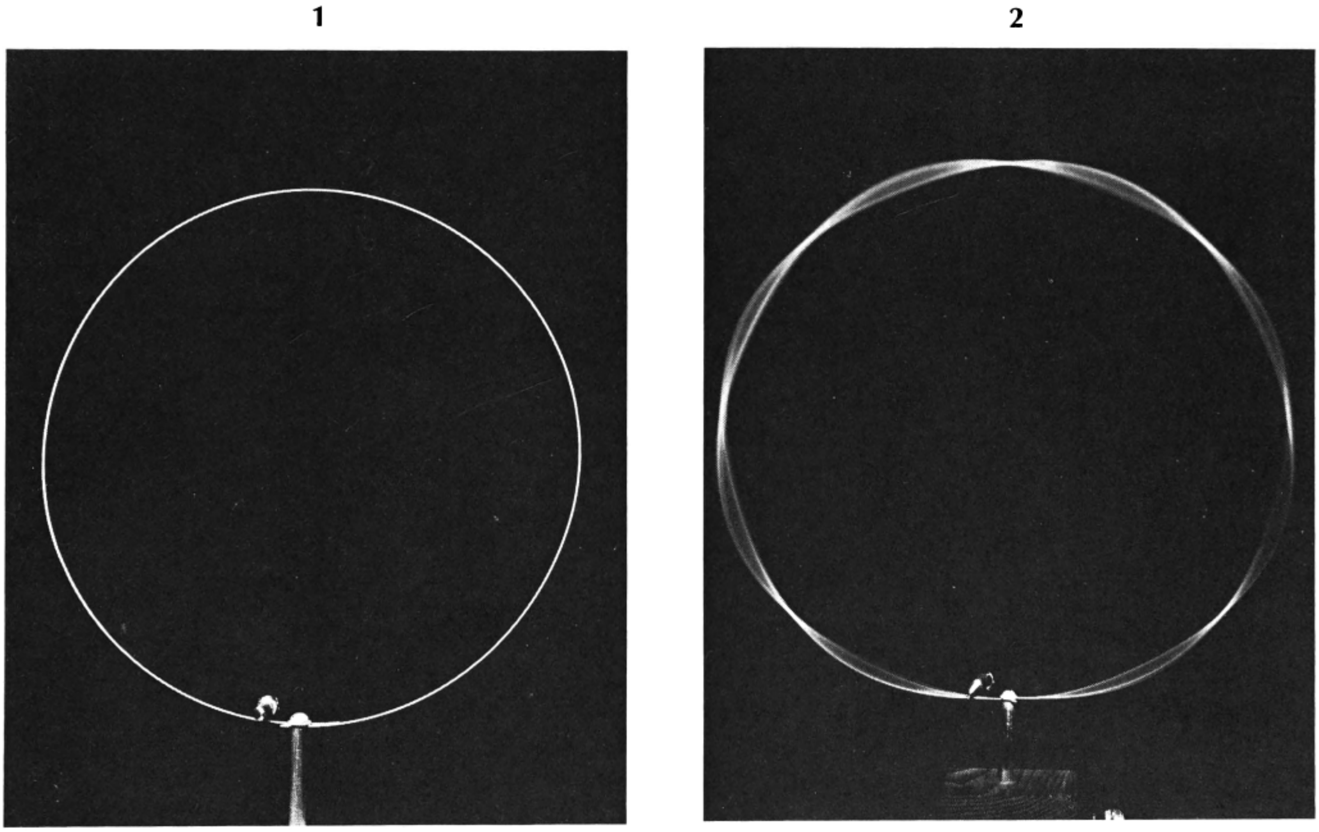
$$E = U_r + E_K = -\frac{kZ}{r} + \frac{kZ}{2r} = -\frac{kZ}{2r}$$

and

$$E = U_r + E_K = -2E_K + E_K = -E_K$$

This is as far as Newtonian mechanics will take us. Neither of these expressions suggests any restriction on the values of the total energy, and therefore they cannot account for the discrete lines in the hydrogen spectrum. Moreover, the hydrogen atoms and many other atoms in their ground state behave like a positive nucleus surrounded by a spherical cloud of negative charge rather than an electron revolving in a circular orbit around a positive nucleus. If, for example, the latter picture were correct, we should be able by some experiment to determine in which plane the electron moves, yet experiments do not show a preferred plane. We must conclude that a spherically symmetric wave pattern is a more appropriate description of the electron in the ground state of the hydrogen atom than a circular orbit. How can we, then, bridge the gap between this description and the orbital description of motion which holds on a larger scale?

We have seen in Section 34 – 2 that the wave nature of particles does not present itself as long



1

2

35 – 1 An illustration of a circular standing wave. Photograph 1 shows a wire loop stationary with no waves on it. Just to the left of the support you can see the shaft of a small electric motor protruding through the black cardboard background. (The motor is behind the cardboard.) The shaft is not circular, so that when it rotates it pushes on the wire, sending waves around the loop. Photograph 2 shows a standing wave with $2\pi r = 4\lambda$.

as all distances which appear in the experiment are large as compared to the de Broglie wavelength. In our case we may expect that Newtonian mechanics will be adequate as long as the circumference of the orbit, $2\pi r \gg \lambda$, the de Broglie wavelength.

Now, from our first expression for the total energy of the electron we have

$$2\pi r = -\frac{\pi k Z}{E}$$

and from the second expression we find

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2mE_k}} = \frac{h}{\sqrt{-2mE}}$$

Thus, Newtonian mechanics will be adequate if

$$\frac{2\pi r}{\lambda} = -\frac{\pi k Z}{E} \cdot \frac{\sqrt{-2mE}}{h} = \frac{\pi k Z}{h}\sqrt{\frac{2m}{-E}} \gg 1$$

We can now try to extend the validity of the relations based on Newtonian mechanics by admitting that even when $2\pi r/\lambda \gg 1$ the standing wave in a somewhat fuzzy circular channel will be a better description than the classical orbit, just as taking into account the very small diffraction of a light beam passing through a wide slit is an improvement on geometrical optics.

For the wave pattern to form a standing wave in this circular channel, the circumference of the channel will have to be an integral multiple of the wavelength;

$$\frac{2\pi r}{\lambda} = n$$

where n is a large integer (Fig. 35 - 1). If n can have only integral values, then E becomes discrete too; we shall denote the energy of the nth level by E_n:

$$\frac{\pi k Z}{h}\sqrt{\frac{2m}{-E_n}} = n$$

By squaring and solving for E_n we find

$$E_n = -\frac{2\pi^2 k^2 m}{h^2} \cdot \frac{Z^2}{n^2} = -R\frac{Z^2}{n^2}$$

The constant

$$R = \frac{2\pi^2 k^2 m}{h^2} = 13.6 \text{ ev}$$

is called the Rydberg constant.*

To find the approximate radius r_n of the channellike wave pattern which corresponds to the nth energy level, we use $E_n = -kZ/2r_n$. Solving for r_n we find

$$r_n = -\frac{kZ}{2E_n} = \frac{h^2}{4\pi^2 km} \cdot \frac{n^2}{Z} = a\frac{n^2}{Z}.$$

---

*Originally the simple formula for the energy states was obtained from the measurements of spectral lines. The spectroscopists Balmer and Rydberg developed it in fitting the spectra into a neat pattern, and what we now call the energy levels were called the term values in their purely spectroscopic analysis. Then in 1913 Niels Bohr showed the connection between the planetary model and the formula $E_n = \frac{-13.6 \text{ ev}}{n^2}$. He was actually able to derive the value 13.6 from the fundamental constants of nature: the Coulomb constant k, Planck's constant h, and the mass m of the electron. We may take Bohr's success as marking the beginning of our present theory of the structure of matter.

Bohr justified his derivation by appealing to a general principle which we now name his Correspondence Principle. Suppose we have a successful theory—a theory developed to deal with the known facts in some region of knowledge. As new knowledge is gained, the theory may show difficulties; it may fail to fit all the new facts. Then a new theory has to be developed to replace the old. But, the new theory has no hope of success unless it agrees with the old theory in all matters where the old theory itself agreed with the facts. In other words, new theory must make the same predictions, in the same form, as old theory in the region where the old was adequate. In this way the structure and content of old theory can provide an essential check on new theory and may even go further and show the direction in which unknown new theory must develop.

Bohr's first use of this Correspondence Principle was in justifying his derivation of the constant for hydrogen energy levels. He imagined the atom excited until it is almost ionized so that the electron moves around the proton in a huge orbit of man-sized dimensions, a meter or so across ($n = 10{,}000$ or so). An orbit like that is the size of a radio antenna or a synchrotron, and we know that in such cases the frequency radiated is the frequency of the motion (up and down the antenna, or around the racetrack). Therefore, the frequency of photons emitted in a hydrogen atom's quantum jumps between levels near ionization must be the same as the frequency of the orbital motion. Equating the photon frequency and the orbital frequency tells us the combination of k, h, and m which must appear in the expression for these energy states and for all other energy states of the hydrogen atom. In this way, Bohr derived his result by relying on an overall principle of fitting together new theory and old. Here, however, we shall derive the result from other considerations.

The constant

$$a = \frac{h^2}{4\pi^2 km} = 0.53 \times 10^{-10} \text{ meter}$$

is called the Bohr radius.

On the basis of our derivation we should trust the relations

$$E_n = -R\frac{Z^2}{n^2} \quad \text{and} \quad r_n = a\frac{n^2}{Z}$$

to agree with the Newtonian predictions only for large values of n. But the expression for the energy levels fits the energies of the observed spectral lines surprisingly well. It holds also for small values of n, as we have seen in Section 33 – 5. In fact, Rydberg found the value of the constant R experimentally before it was calculated by reasoning similar to that we presented here. We may not be so lucky in applying the expression for r_n for small values of n, since we know that even the fuzzy orbit picture breaks down as we approach the ground state of the atom. Nevertheless, for $n = 1$ and $Z = 1$ we get $r_1 = 0.53 \times 10^{-10}$ meter, which is roughly the size of a hydrogen atom. In the following sections we shall try to extend the expression for E_n and r_n to more complicated atoms, where they will be helpful to us in understanding their structure.

35–2 THE HELIUM ION AND THE HELIUM ATOM

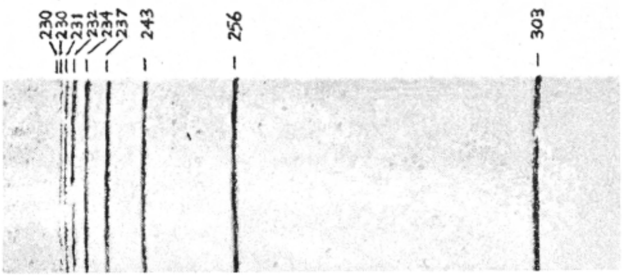
The helium atom has two electrons and its nucleus carries two elementary positive charges, which makes a system of three interacting particles; there is no longer a simple central force field as in the hydrogen atom. In general this results in a rather complex structure of energy levels. But if one of the electrons is removed, as may happen in a collision with another atom or an energetic electron, we are left with a positive ion: a single electron bound in the field of a nucleus $Z = 2$. To the ion we can apply the wave theory of hydrogenlike atoms.

Since now $Z = 2$, we expect the following energy levels:

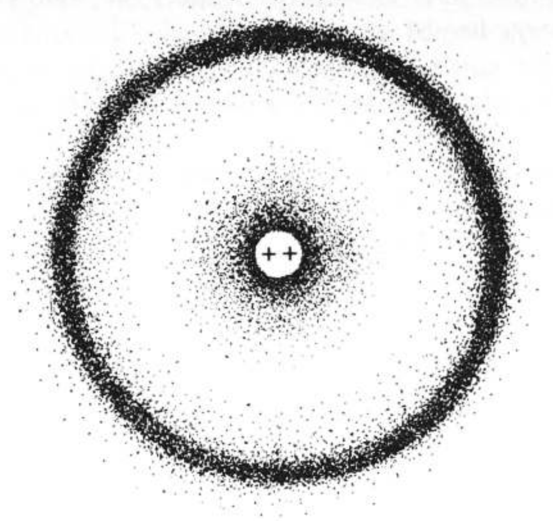
$$E_n = -13.6 \times \frac{2^2}{n^2} \text{ ev} = -\frac{54.4}{n^2} \text{ ev}$$

This is the same arrangement of levels as in hydrogen, only scaled up by a factor $Z^2 = 4$. Thus, the energy of the photons emitted in any given jump also will have four times the energy of those emitted in the corresponding jumps in the hydrogen atom. Since the wavelength of a photon is inversely proportional to its energy, the lines will be shifted farther into the ultraviolet, as compared with the corresponding hydrogen lines. This is amply confirmed by experiment (Fig. 35 – 2).

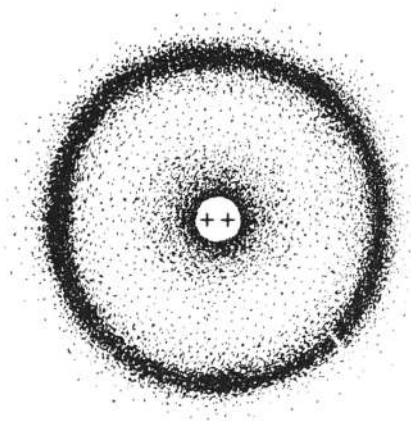
Suppose now that a helium ion gets hold of a second electron. When this happens, it is most likely that the second electron will first be in an "orbit" corresponding to a large value of n. A schematic drawing is shown in Fig. 35 – 3. The electron close to the nucleus screens off one of the two positive charges on the nucleus, and the electron in the outer orbit feels the net effect of a single positive charge. We expect the energy levels corresponding to the outer orbits to be almost the same as those of atomic hydrogen. Therefore, as the electron jumps down to lower and lower levels, it will emit photons of very nearly the same energy as those emitted by hydrogen atoms. But when the electron gets close



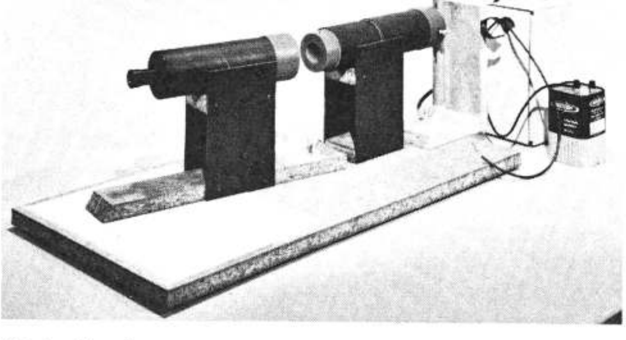
35–2 The spectrum of singly ionized helium with the wavelengths marked in angstroms. These lines correspond to the Lyman series of hydrogen modified for a nucleus with 2 elementary charges. Notice that they are in the extremely short ultraviolet range. (From: P. G. Kruger, "Arc and Spark Spectrum of He," Physical Review, Vol. 36, Series 2, July-December 1930.)



35–3 Sketch of the cross-sectional view of the helium atom in a nearly ionized state when n is large. (Not drawn to scale.)



35–4 Sketch of the cross-sectional view of the low excited state of helium. (Not drawn to scale.)



35–5 Simple spectrometer.

enough to the nucleus, its standing wave pattern will begin to overlap with that of the inner electron (Fig. 35 – 4). The electric field can no longer be regarded as produced by a point charge.

In fact, the force field is rather complicated, being composed of the force due to the nucleus and that due to the other electron. These forces are comparable in strength, since the nucleus has only twice the charge of the electron. We cannot, therefore, hope in general to describe the wave patterns of the two electrons in terms of the wave patterns of the hydrogen atom. Only when the two fields, that of the nucleus and that of one electron, add up in such a way as to make up a field similar to an inverse-square field will the hydrogen atom model be of use. This is indeed the case if the wave pattern of one electron is far outside the wave pattern of the other. Our approximation will, therefore, consist of taking into account the decrease in screening of the electric field of the nucleus as the wave patterns of the electrons overlap but ignoring the distortion of the inverse-square field. If we use this approximation we shall expect to find hydrogen-like energy levels with a value of Z between 1 and 2. We shall refer to such values of Z as Z_{eff} (Z-effective). Far away from the nucleus the second electron will move in a field where $Z_{\text{eff}} = 1$. When the electron is as close as it can get, Z_{eff} will be at most the full charge of the nucleus, that is, $Z_{\text{eff}} = 2$, but it is likely to be less. We can calculate Z_{eff} for the ground state from the observed ionization energy of helium, which is 24.6 ev. With $n = 1$ we find

$$-24.6 = -13.6\,\frac{Z^2_{\text{eff}}}{1^2} \text{ or } Z_{\text{eff}} = \sqrt{1.81} = 1.35$$

The excited levels have different effective charges, but for most of them $Z_{\text{eff}} = 1$, that is, the inner electron screens off one of the positive charges on the nucleus. You can convince yourself of the validity of this approximation by actually analyzing the visible spectrum of helium in the laboratory. The simple spectrometer shown in Fig. 35 – 5 will suffice.

35–3 THE LITHIUM ATOM

Now let us turn our attention to the lithium atom. Its nuclear charge is three units, and the neutral atom therefore has three electrons surrounding it. When we excite lithium atoms by means of an electric arc and by generating high-energy sparks, we obtain a spectrum of lines which extends over a wide range of wavelengths, from the far ultraviolet (10^2 Å) to the far infrared (7×10^4 Å). The middle region of the spectrum (1000 – 10,000 Å) contains many lines, and it is difficult to find any order there when we are armed only with an understanding of the hydrogen atom. But at the two extremes the structure is simpler. Let us start with the high-energy end of the spectrum.

Table 1 shows the wavelengths and energies of the most energetic lines.

What can we learn about the energy levels responsible for these lines? The most energetic line has an energy of close to 9R, the binding energy of an electron in the ground state in the field of a nucleus with $Z = 3$, which is indeed the lithium nucleus. This suggests that the lines recorded in Table 1 are due to transitions in the doubly ionized lithium atom. To strengthen this interpretation we have to check up on the numbers in the third column of the table and see whether they can be represented as differences of energy values which have the form

$$E_n - E_1 = 9\left(\frac{1}{1^2} - \frac{1}{n^2}\right)$$

where $n = 2, 3, 4$, etc. To start with, consider the case $n = 2$:

$$\frac{1}{1^2} - \frac{1}{2^2} = \frac{3}{4}, \text{ and } 9 \times \frac{3}{4} = 6.75;$$

this is exactly the energy of the second line in the table. You can easily convince yourself that the following lines correspond to transitions from $n = 3, 4$, and 5 to the ground state of doubly ionized lithium. The 199.3 Å does not fit into this scheme; we shall come back to it shortly.

The infrared end of the spectrum of lithium is presented in Table 2. Here we find that the simple hydrogen levels $-1/n^2$ (in rydbergs) give reasonably good values for the photon energies of the lines. To take the first line in the table as an example, its photon energy is 0.0715 rydbergs, in good agreement with the energy-level difference of $\frac{1}{3^2} - \frac{1}{5^2} = \frac{1}{9} - \frac{1}{25} = 0.0710$. As you can find for yourself, the other lines show even closer correspondence.

TABLE 1 *The Far Ultraviolet Portion of the Lithium Spectrum*

λ(Å)	E (EV)	E (RYDBERGS)
199.3	62.3	4.57
135.0	91.8	6.75
113.9	108.9	8.02
108.0	114.7	8.40
105.5	117.5	8.65
104.2	118.8	8.73

TABLE 2 *The Infrared Portion of the Lithium Spectrum*

λ(Å)	E (EV)	E (RYDBERGS)
12,782.2	0.970	0.0715
18,697.0	0.663	0.0488
40,475	0.306	0.0225
74,360	0.167	0.0123

These lines seem to be produced by an electron moving in an inverse-square field about a central charge with $Z = 1$. As we did in the case of helium, we can think of the central charge as made up of the nucleus closely screened by the two other electrons of the lithium atom. In a field of this kind, the maximum of the wave pattern for the outer electron lies at a distance of n^2a from the central charge. For these far infrared lines, this is at distances of 9 to 36 Bohr radii. This is far enough to ensure no overlap with the wave pattern of the screening electrons.

What about the ground state of the lithium atom? If the screening of the nucleus remains complete, we expect the third electron to have the binding energy of hydrogen, 13.6 electron volts or 1 rydberg. If there is some penetration of the wave patterns, we expect Z_{eff} to be greater than 1, and the binding energy to be larger than 1 rydberg, as it is with helium.

Actually the binding energy of the first electron is lower, 5.39 ev or 0.396 R. Can we account for this surprisingly low binding energy?

Let us look again at our formula for the binding energy:

$$E_b = -E_n = RZ^2_{\text{eff}}/n^2.$$

Z_{eff} cannot be lower than 1. If our approximation is valid, the only remaining possibility is then that the standing wave pattern of the third electron corresponds to a value of n greater than 1. The experimental value of the binding energy indicates strongly that the third electron has gone into a state with $n = 2$. Using the value $n = 2$ in our formula yields $Z_{\text{eff}} = \sqrt{4 \times 5.39/13.6} = 1.26$, a value which appears quite reasonable. Although the third electron has more energy in the $n = 2$ level, it somehow fails to emit a photon and join the other two electrons which are already in the $n = 1$ level. Is the third electron in lithium an exception? Will further electrons all congregate in the $n = 2$ level, will they settle in levels of higher values of n, or will they fall back to the $n = 1$ level? We shall find the answer shortly, but first let us put the pieces of the lithium story together.

From the high-energy spectrum of lithium (Table 1) we know that the nucleus of the lithium atom has $Z = 3$, and that at least one electron can enter the energy state for which $n = 1$. From the low-energy spectrum and from the binding energy, we know that one electron can enter energy states for which $n = 2, 3, 4, 5, 6$, and that it is more or less completely screened from the nucleus by two electrons in between. If we ionize a lithium atom, the binding energy of the second electron comes out to be 75.6 ev, and if we set $-75.6 = -13.6\ Z^2_{\text{eff}}/1^2$, we find $Z_{\text{eff}} = 2.36$, which indicates a screening by the remaining electron to about the same extent as in the ground state of the helium atom. In addition, we can find in the ultraviolet spectrum of lithium a set of lines which resembles on a different scale the spectrum of the helium atom, and which we can assign to the lithium ion with two of its electrons still attached. The line 199.3 Å belongs to this set. This adds up to good evidence that there are two electrons in the energy levels for which $n = 1$.

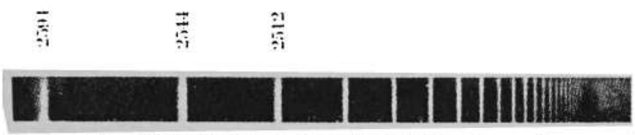
The maxima of the wave patterns for the different electrons lie at considerably different distances from the nucleus. Using the values of Z_{eff} which we have calculated from their binding energy, we find for the inner electrons $r_1 = (1^2/2.36)\ a = 0.42a$, and for the outer, $r_2 = (2^2/1.26)\ a = 3.2a$. Of course, these distances indicate only where these electrons are most likely to be, and certainly there is overlapping between the outward spread of the inner wave pattern and the inward reach of the outer one. Nevertheless it makes sense to think of these electrons as forming more or less separate spherical shells about the nucleus, with two tightly bound electrons in the inner shell and one more loosely bound in the outer.

Let us see whether we can find evidence for such a shell structure in other atoms.

35–4 ABSORPTION SPECTRA

We have seen in Section 33 – 4 that if we send white light through a gas, most of the photons pass through but a few are absorbed, and their absence can be noted in the spectrum of the light by a series of dark gaps across it. A photon is absorbed if it can transfer energy to an atom of the gas, and this happens when the photon's energy is exactly enough to shift an electron from the ground state to a state of higher energy, $h\nu = E_{\text{final}} - E_{\text{initial}}$. An absorption spectrum of this type is shown in Fig. 35 – 6. Each of the lines in it measures the energy difference between the ground state and one of the excited states, and they crowd together at the high-energy end as the excited states fall closer together for larger and larger values of n.

The limit of the absorption spectrum is reached when a photon has enough energy to remove the electron from the atom entirely, that is, to ionize the atom. Ionization by light is something we have met before. We called it the photoelectric effect, and we remember that in the photoelectric effect the electron can leave the atom with any kinetic energy whatever. That is to say that any photon which has more than the minimum ionization energy of the atom can be absorbed. The elimination of all these photons from the spectrum leaves a wide, blurred band just beyond the limit of the sharp-line series. This band shows at the right end of Fig. 35 – 6, if you remember that the illustration represents a negative in which light and dark are reversed.



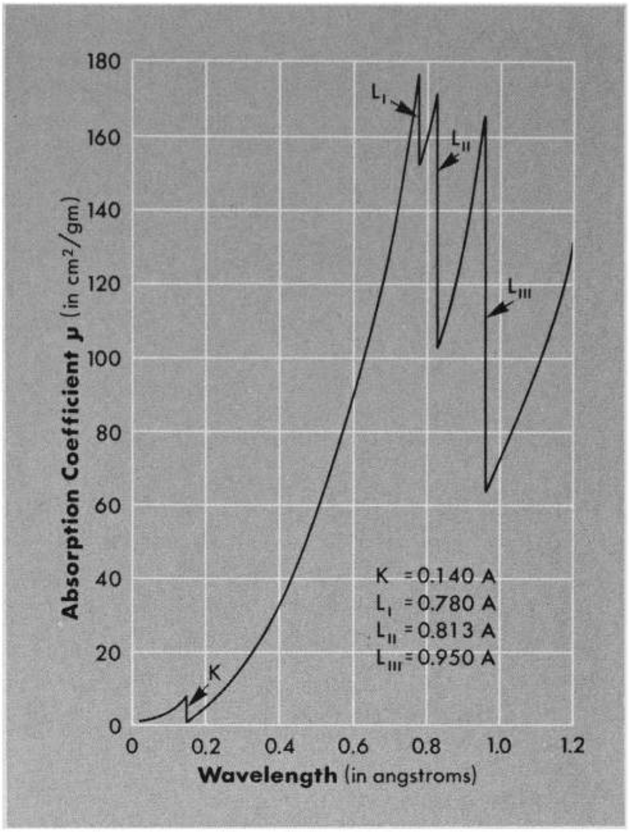
35–6 The short-wavelength region of the absorption spectrum of sodium. Since this is a reproduction of a photographic negative, the wave-lengths absorbed appear as light lines on a continuous dark background. (From H. Kuhn, "Uber-Spektren von unecht gebundenen Molekulen (Polarisationsmolekulen) K_2, Na_2, Cs_2, and Verbreiterung von Absorptionslinien," Zeitschrift fur Physik, Vol. 76, 1932.)

X rays are more penetrating than visible or ultraviolet light, but they are absorbed by dense materials. If we send X rays of all possible wavelengths through a thin layer of material and take the spectrum of what emerges on the farther side, we find that most photons come through. The absorption is limited to a few wide bands which fade out gradually toward high energies and are cut off by a sharp edge on the low-energy side (Fig. 35 – 7).

The fading off of the band suggests that this is another ionization process, but one without a series of absorption lines leading up to it. The electrons involved in X-ray absorption seem to have no nearby excited states to shift to. This would be the case if they were inner electrons which had to be taken entirely outside the atom in a single move. We could then understand the sharp edge which cuts off the low-energy side of the band. No photon could be absorbed which had less than the ionization energy, and any photon could be absorbed which had greater than that minimum energy.

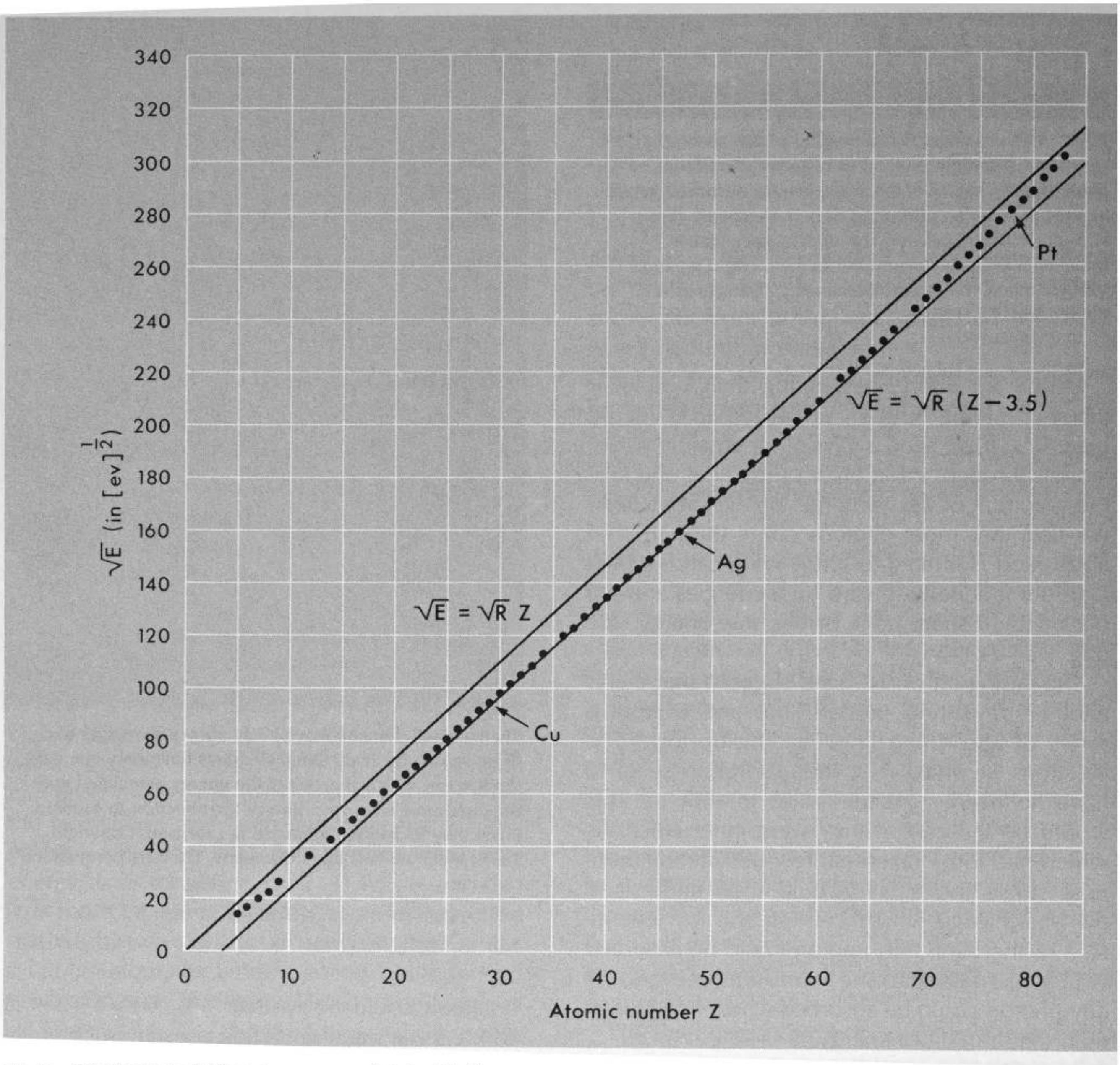
Suppose we test this idea by examining the photon energy for the absorption edges of the different elements. This should be the binding energy for an inner electron, and if the ideas we developed for hydrogen are still useful with more complex atoms, we might express it as $E_b = RZ^2_{eff}/n^2$. Most elements have more than one X-ray absorption edge. The one which represents the highest photon energy should belong to the innermost electrons, quite possibly with $n = 1$ and Z_{eff} very close to the full nuclear charge.

A convenient test would be to make a graph in which $\sqrt{E_b}$ was plotted against Z. We might expect this graph to be a straight line whose equation could be written $\sqrt{E_b} = \sqrt{R}\,(Z - s)$, where s is a small screening constant such as we have met already with helium and lithium.



35–7 Most elements have X-ray absorption edges like these shown for lead. Some elements have only one edge, the K edge, which represents the energy associated with the innermost electrons. (From: Introduction to Modern Physics by Richtmyer, Kennard & Lauritsen. Copyright 1955. McGraw-Hill Book Company. Used by permission.)

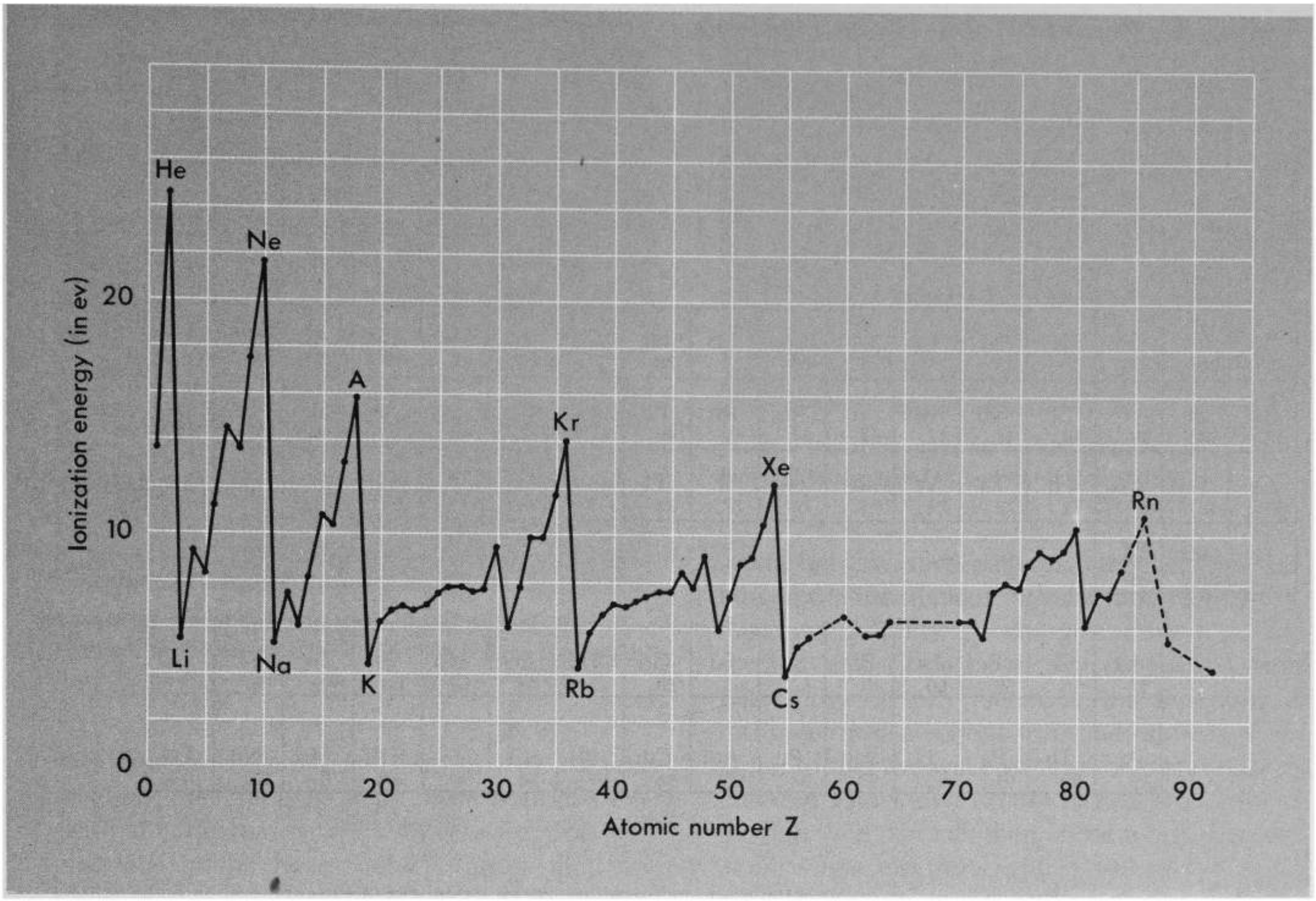
Figure 35 – 8 shows the graph. The suggestion was made in Section 26 – 4 that the nuclear charge for each element might be the same as its atomic number—that is, its serial number in the early chemical list—where the elements have been arranged in the order of their atomic masses. We have used that value of Z here. For three elements we have independent confirmation. Chadwick has used alpha-particle scattering to make careful measurements of the nuclear charges of copper (No. 29), silver (No. 47), and platinum (No. 78). His values of 29.3, 46.3, and 77.4 differ from the atomic numbers by less than 2 percent.



35–8 E is the highest X-ray energy associated with the absorption edges for a particular element. $\sqrt{E}$ is plotted against the atomic number of the element. All the data points lie between the two straight lines representing $\sqrt{E} \propto Z$ and $\sqrt{E} \propto (Z - 3.5)$.

The upper line represents $\sqrt{R}\,Z$, the lower, $\sqrt{R}\,(Z - 3.5)$, and we see that all the points fall between the two. It is clear that our simple hydrogenlike calculation of energy is not quite correct but, as the graph shows, even in quite complex atoms it is still a good approximation. We can guess at the reasons why it is not perfect. In an atom with 30 or 40 electrons, there must be a good deal of overlap in the wave patterns and there is no longer an inverse-square field near the center.

Nevertheless, there are two conclusions we can draw from Fig. 35 – 8 with a good deal of assurance. One is that the electrons which are involved in these high-energy X-ray absorption edges are the innermost electrons of the atom, in energy states for which $n = 1$, and moving in substantially the full field of the nuclear charge.



35–9 Ionization energy of the outermost electron of each of the elements.

The other is that the nuclear charge increases by uniform steps as we go from element to element along the chemical list. While we cannot say with certainty that the nuclear charge is exactly equal to the atomic number for any one of the elements, there is no other value which will accommodate the whole group of them.

35–5 IONIZATION ENERGIES

Figure 35 – 8 is a graph of the square root of the ionization energies for the innermost electrons of the elements, that is, of the energy needed to remove one of these electrons entirely from each atom. Figure 35 – 9 is a graph of the ionization energy for the outermost electrons. The difference between the two is striking. For the outer electron, there is no smooth progression. The ionization energy sometimes increases and sometimes decreases as we go from element to element. In place of a steady increase as we go to higher atomic numbers, we find in spite of the fluctuations a slow, general decrease. The binding energy of the outermost electron evidently depends on factors other than the charge on the nucleus.

In Fig. 35 – 9 we notice especially the sharp peaks (He, Ne, Ar, Kr, Xe) and the sharp dips which directly follow them (Li, Na, K, Rb, Cs). The peaks belong to the noble gases helium, neon, argon, krypton, and xenon. The dips belong to the alkali metals which react with great vigor, oxidizing rapidly in the air and decomposing water when they dissolve in it. Between each alkali metal and the next noble gas the graph climbs irregularly.

There is a periodic change here which is very

TABLE 3 *Periodic Table of the Elements*

I	II											III	IV	V	VI	VII	O
H 1																	He 2
Li 3	Be 4											B 5	C 6	N 7	O 8	F 9	Ne 10
Na 11	Mg 12											Al 13	Si 14	P 15	S 16	Cl 17	Ar 18
K 19	Ca 20	Sc 21	Ti 22	V 23	Cr 24	Mn 25	Fe 26	Co 27	Ni 28	Cu 29	Zn 30	Ga 31	Ge 32	As 33	Se 34	Br 35	Kr 36
Rb 37	Sr 38	Y 39	Zr 40	Nb 41	Mo 42	Tc 43	Ru 44	Rh 45	Pd 46	Ag 47	Cd 48	In 49	Sn 50	Sb 51	Te 52	I 53	Xe 54
Cs 55	Ba 56	See Below 57-71	Hf 72	Ta 73	W 74	Re 75	Os 76	Ir 77	Pt 78	Au 79	Hg 80	Tl 81	Pb 82	Bi 83	Po 84	At 85	Rn 86
Fr 87	Ra 88	See Below 89-															

La 57	Ce 58	Pr 59	Nd 60	Pm 61	Sm 62	Eu 63	Gd 64	Tb 65	Dy 66	Ho 67	Er 68	Tm 69	Yb 70	Lu 71
Ac 89	Th 90	Pa 91	U 92	Np 93	Pu 94	Am 95	Cm 96	Bk 97	Cf 98	Es 99	Fm 100	Md 101	No 102	Lw 103

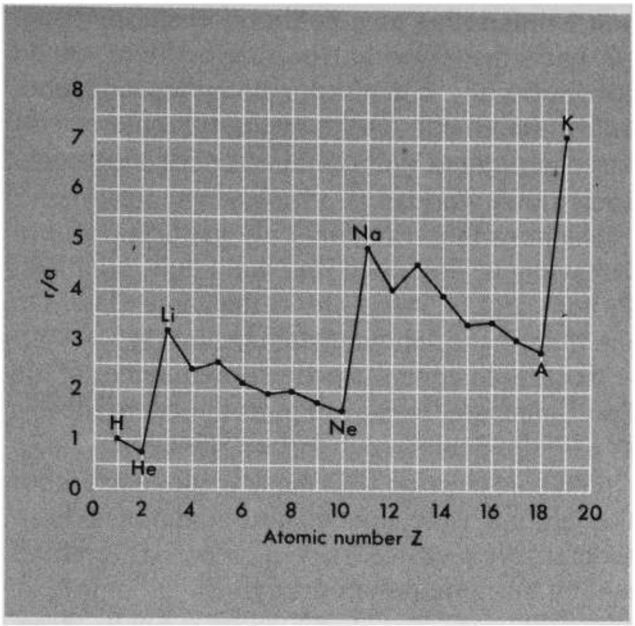
much like the periodic changes in the chemical list of the elements (Table 3). It repeats, in fact, at exactly the same places. The alkali metals occupy the first column of the periodic table into which the list is broken up; the noble gases, the last column. These ionization energies are quite evidently related to the chemical properties of the atoms.

We have seen that the drop in ionization energy (or binding energy) between helium and lithium corresponds to an increase in n by 1. Let us see whether the same explanation will work for the drop between neon and sodium. The binding energy is RZ^2_{eff}/n^2. Then for neon we have $21.6 = 13.6\,Z^2_{\text{eff}}/2^2$, so that $Z_{\text{eff}} = \sqrt{6.35}\,n^2 = 2.52$, and for sodium, $5.14 = 13.6\,Z^2_{\text{eff}}/3^2$, so that $Z_{\text{eff}} = \sqrt{3.41} = 1.85$. These are larger than the effective charges we had for helium (1.35) and lithium (1.26), but we remember that the nuclear charges are now 10 and 11 instead of 2 and 3. Nevertheless, it is encouraging that the effective charge has approximately the same value in each case. With very nearly the same number of electrons in the two atoms, we should expect about the same degree of screening.

35–6 THE SIZE OF ATOMS[*]

Now let us estimate the size of the atoms of neon and sodium. When we think in terms of the wave patterns of their electrons, we see that we cannot speak as though an atom had definite boundaries to measure between. Nevertheless we can get some idea of size by using our expression for the radius at which the maximum of the wave pattern lies, $r_n = a\,n^2/Z_{\text{eff}}$. This gives us for neon, $r_{\text{Ne}} = a \cdot 2^2/2.52 = 1.59a$, and for sodium $r_{\text{Na}} = a \cdot 3^2/1.85 = 4.85a$. The calculated radii for the atoms from hydrogen to potassium are shown in Fig. 35 – 10.

We should like to check up on this by measuring the sizes of the two atoms, but we shall need care in choosing a tool with which to measure. We need something delicate enough to "feel" the outer electrons of the atoms without disturbing them. In fact, the most suitable probe probably would be another electron. Experiments have been made of sending a beam of electrons through neon gas and through sodium vapor and observing how much they are scattered. The atomic radii worked out from two of these



35–10 Graph of the ratio r/a vs. Z for the first 19 elements in the periodic table. The values were computed from Z_{eff} as determined from the ionization energies.

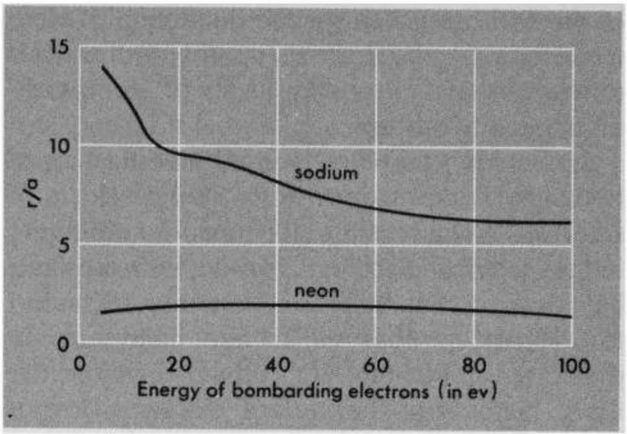
experiments are plotted in Fig. 35 – 11. As you can see, the "radius" of an atom from this experiment decreases with the energy of the electrons in the beam, which means that the result of the measurement does depend on the measuring tool. For all that, the sodium atom is clearly bigger than the neon atom.

35–7 ELECTRON SHELLS

We have seen that the helium atom has two electrons with $n = 1$, and the lithium atom has also two electrons with $n = 1$ but the third with $n = 2$. The spectra of the elements that follow lithium have been studied with violent sparks which strip electrons away, and we know that as far up the chemical list as sodium and magnesium, no atom has more than two electrons with $n = 1$.

The ten electrons of the neon atom would then fall into two groups, two of them with $n = 1$ and eight of them with $n = 2$. The eleventh electron of sodium seems to take $n = 3$. Again the study of spectra confirms this and tells us also that from sodium at least as far up the list as calcium, no element has more than eight electrons with $n = 2$.

We cannot say where any one electron is in an atom, but the wave patterns tell us at what distance from the nucleus they are most likely to be. From our graph of the energies of the X-ray absorption edges, we know that even in an atom with thirty or forty electrons, the screening of the innermost electrons from the nucleus is not large, so that the wave patterns of electrons with different values of n do not have a great deal of overlap. We can talk then about a region in the atom where electrons with $n = 1$ or $n = 2$ are most likely to be found, and we can speak of the *electron shell* which occupies this region.



35–11 The value of r/a for sodium and neon as a function of the speed of bombarding electrons.

The absorption edges we have plotted on the graph of Fig. 35 – 8 tell us about electrons which have been removed from the innermost shell of each atom. One interesting thing about the X-ray absorption bands is that there is no series of lines leading up to them, but only the sharp edge to mark the energy at which the absorption of photons begins. There are no energy levels open to electrons from an inner shell until they are entirely outside the atom.

Thus from X-ray spectra, from ultraviolet spectra, and from binding energies, we find reasons to think that there is a limit to the number of electrons which can occupy each shell—two for the innermost, eight for the second and third, eighteen for the fourth and fifth. When a shell has all the electrons it can take, we often speak of it as "closed."

35–8 THE PAULI PRINCIPLE

The existence of electron shells is a phenomenon which we could not predict from the study of the hydrogen atom or the diffraction of electrons by a crystal. It shows up only in the study of the spectra and energy levels of atoms containing many electrons.

To be sure, the experimental investigation of hydrogen is never done with a single electron: there are always many electrons—for example, in the electrical discharge tube which serves as a light source. But these electrons are all bound to different protons which are many atomic diameters apart from each other. Therefore the wave patterns of the various electrons do not overlap, and we study many isolated systems, each containing a single electron.

The same is true for the diffraction of electrons where a beam containing many electrons is used. We can perform the experiment with very weak beams so that only a few electrons are in the apparatus at any moment, and these few are quite far apart, to assure that their wave patterns do not overlap.

The situation is quite different when we come to an atom with many electrons. Here all the electrons are bound by the electric field of the same nucleus, so their patterns must overlap. Moreover, it is plausible that in the ground state all the electrons have the same standing wave pattern. This picture is consistent with experimental results for helium but not for lithium (Section 35 - 3). Only two electrons bound to a nucleus by the Coulomb field can have a wave pattern characterized by the quantum number $n = 1$. This is a fundamental property of electrons, just as fundamental as their charge or mass.

If no more than two electrons can have the same wave pattern around the same nucleus, why are there more than two electrons in the $n = 2$ shell? The answer is that to describe a wave pattern in three-dimensional space requires three numbers and not just one, as we shall see in the next section. The fact that no more than two electrons can have the same wave pattern is known as the Pauli principle and is not limited to the Coulomb field; it holds regardless of the nature of the field in which the electrons move.*

It is interesting to note that if electrons could all have the same wave pattern, there would hardly be any chemistry. All atoms, independent of the charge on the nucleus, would resemble each other. We would not exist to observe such a dull world.

Electrons are not the only particles whose behavior is described by the Pauli principle. Protons and neutrons are in the same class. On the other hand, photons are not.

In order to appreciate the consequences of the Pauli principle, let us return to the energy levels of particles contained in a "box" (Section 34 - 7).

For one electron in a box, we found that for the ground state, the quantum numbers have their smallest values ($n_x = n_y = n_z = 1$) and the system has a minimum energy

$$E_0 = 3h^2/8md^2$$

Now suppose we add electrons to the box, neglecting the electrostatic repulsion between them.

The Pauli principle implies that no more than two electrons in the same box may have the same standing wave pattern, so that while a second electron could also have $n_x = n_y = n_z = 1$, a third electron would have to have a different pattern, with at least one of the quantum numbers being different from 1. As a consequence, the third electron could not have the same energy as the first two, but must rather have a greater energy than either of the first two electrons, corresponding to n values (1, 1, 2) or (1, 2, 1) or 2, 1, 1). This corresponds qualitatively to what we have seen in the behavior of the third electron in lithium.

There can be two electrons with $n_x = 1$, $n_y = 1$, $n_z = 2$; two with $n_x = 1$, $n_y = 2$, $n_z = 1$; and two with $n_x = 2$, $n_y = 1$, $n_z = 1$. Again ignoring any interaction among them, each of the six of them will have an energy

$$E_1 = \frac{(1^2 + 1^2 + 2^2)h^2}{8md^2} = \frac{6h^2}{8md^2}$$

A ninth electron added into the box will have a different set of quantum numbers and a still higher energy.

Note that the finite size of the atom, together with the Pauli principle, accounts for the fact that not all the electrons in an atom are bound equally strongly: those in the higher energy levels will require less energy to be ejected from the atom than those in the lower levels. Of course, we cannot hope to account for detailed structure of atomic energy levels with such a simplified model which ignores the interaction among the electrons, attributes the same size of box to all electrons independent of their energy, and approximates a basically spherical atom by a cubical box.

---

*There is another property of the electron which has nothing to do with the probability wave pattern in space. It is its intrinsic angular momentum or spin. When two electrons have the same wave pattern, they always have opposite spins. Thus the usual formulation of the Pauli principle says that only one electron can have a given wave pattern and spin.

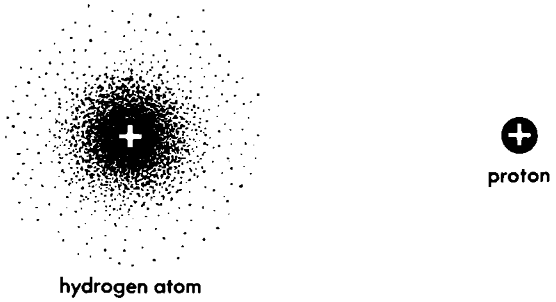
35-9 CHEMICAL BINDING

We have seen in Section 35 - 5 that the ionization energies and the chemical properties of atoms show similar periodic dependence on the atomic number. Look, for example, at the columns of the noble gases He, Ne, Ar, Kr, and Xe in Table 3. These same elements are at the peaks of the curve in Fig. 35 - 9. Their ionization energies are considerably higher than those of their neighbors. Even the outermost electrons are relatively tightly bound in these atoms, and this accounts for their sparse chemistry. Notice, though, that the ionization energies of xenon and krypton are less than those of atoms of several elements which do form compounds. Recently compounds of xenon and krypton with fluorine and oxygen have been prepared.

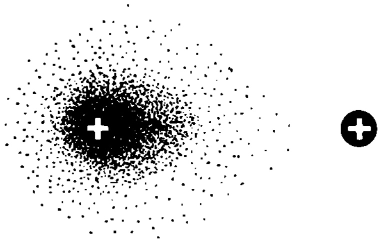
It is quite plausible to suspect that the chemical properties of atoms are primarily governed by their outermost electrons, the electrons in the shell with the largest value of n. Detailed predictions about the formation of molecules require a detailed analysis. We shall not follow so ambitious a program here, but rather limit ourselves to a qualitative discussion of the simplest combination of atoms, the molecular hydrogen ion, which is formed when a hydrogen molecule loses one electron or when a hydrogen atom combines with a proton; it consists therefore of two protons and one electron.

The main question which presents itself is why two protons and one electron form a bound system. The key to the answer to this question is the recognition that a bound electron is confined to a finite region of space, and the probability of finding it there is described in terms of a standing wave pattern. When there is an increase in the region in which the electron is bound, the wavelength increases too, and therefore the kinetic energy of the electron decreases.

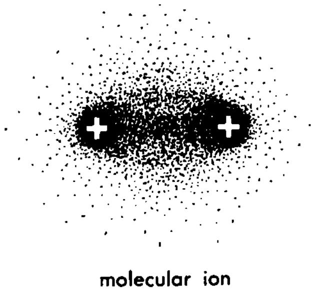
Consider, now, a hydrogen atom and a proton far away (Fig. 35 - 12). Since the atom is neutral, there is no force between the proton and the atom. Suppose we bring the proton closer to the atom (Fig. 35 - 13). Since the charge of the electron is "spread" spherically around the proton, part of it is now considerably closer to the proton on the right than to the proton on the left. The pattern will get a little distorted, and a net attraction results. This attraction will continue until the two protons are so close that a further reduction of the distance between them would increase their mutual repulsion so much as to make it greater than the attraction. If we have a net attraction at larger distances and a net repulsion at smaller distances, then in between there is an equilibrium position with a minimum of potential energy of the system (Fig. 35 - 14).



35-12 The proton is so far away from the neutral hydrogen atom that there is no attraction between them. The spherical volume shown about the nucleus of the atom represents the space where the probability of finding the electron is appreciable.



35-13 When the proton is moved closer to the atom, it distorts the standing wave pattern of the electron around the nucleus.



35-14 The approximate shape of the wave pattern of the electron in the molecular hydrogen ion.

What about the kinetic energy? If the "size" of the molecular ion were about the same as the "size" of the hydrogen atoms, then the kinetic energy of the electron in both cases would be about the same. Actually, the molecular ion is larger than the atom. This increase in the size reduces the kinetic energy of the electron and further lowers the total energy of the system as compared with the situation shown in Fig. 35 – 12.

All chemical bindings have these two ingredients: a lowering of potential energy due to electrostatic attraction, and a lowering of kinetic energy due to an increase of the region over which the outer electron or electrons are "smeared." When there are many electrons involved, the Pauli principle has to be taken into account. The existence of molecules and chemical binding, just like the very existence of the hydrogen atoms, cannot be explained by Newtonian physics. The wave aspects of particles are at the root of the matter.

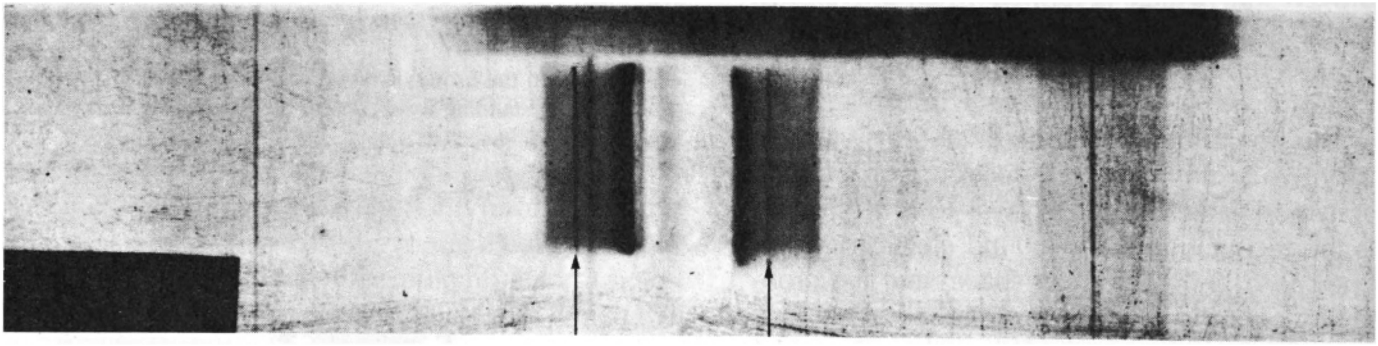
35–10 THE DEUTERON

We have seen that the general behavior of molecules and atoms can be understood by the use of quantum theory and the idea that they are made up of electrons and nuclei with Coulomb forces between them. Now, what about the nuclei themselves? We know that they are very small (of the order of 10^{-14} m; see Section 26 – 4) and that they carry most of the atomic mass. They are made up of neutrons and protons, which are bound together. What is it that holds the protons and neutrons together? Can we understand nuclei even qualitatively, using our quantum ideas?

Let us start with the simplest nucleus containing both neutrons and protons, namely one of each. This nucleus is called a deuteron. In a way it is the nuclear analogue of the hydrogen atom consisting of a proton and an electron. However, the binding force cannot be the electric force, since the neutron carries no electric charge. We shall now compare the deuteron and the hydrogen atom in some detail to see the significance of the difference.

A deuteron is formed when a proton captures a slow neutron. This capture process is in many ways very similar to the capture of an electron by a proton to form a hydrogen atom, but it is on an entirely different energy scale. (See Section 32 – 7.) When slow neutrons are captured by protons, all the emitted photons have an energy close to 2.2×10^6 ev as compared with the maximum energy of 13.6 ev carried off by photons when electrons are captured by protons to form a hydrogen atom. Actually this energy has been measured to four significant figures, using crystal diffraction gratings like those used to measure the wavelengths of X rays. (Fig. 35 – 15). The reason we get photons with a nearly unique energy in the capture process is that the slow neutrons used actually have negligible kinetic energy compared with the energy of the photons.

A very remarkable property of this capture process is that the nuclear spectrum has only one line. In the capture process of very slow neutrons we see only one photon energy—



35–15 A photographic plate showing the gamma-ray line resulting from the capture of neutrons by protons. The gamma ray, diffracted by a crystal grating, produced a first order line on each side of the central maximum. The lines are indicated by arrows. The central maximum itself was blocked out. The pair of outer lines are calibration lines, due to first-order diffraction of 0.511 Mev photons from positron-electron annihilation. As with usual diffraction effects, wavelength is proportional to distance along the plate. (From: Kazi, Rasmussen, and Mark, Physical Review, 123, 1310, 1961.)

2.2 Mev. This is quite different from the case of the hydrogen *atom*, where the spectrum contains very many lines. It appears that the deuteron has only one bound state—the ground state.

To investigate this point further, let us, consider the analogue of the ionization experiments on hydrogen. In the case of the deuteron we need photons in the Mev range, and these were obtained in the experiment by means of a large electrostatic accelerator, which produced a beam of monoenergetic electrons. The electrons were stopped in a target, producing photons with a maximum energy equal to the energy of the electrons in the beam—a sort of giant X-ray tube. The photons were then allowed to strike a deuterium target, and the number of neutrons produced was measured as a function of the energy of the electron beam. Neutrons were produced in abundance only when the incident energy was greater than 2.2 Mev. To pull the deuteron apart takes the same energy as we get when one is formed from a neutron and a proton. This, then, confirms the idea that the deuteron has only one bound state with binding energy 2.2 Mev.

Here we have a real quantum effect, but a very different one from the case of the hydrogen atom. Table 4 summarizes the contrast between the two. The form of the energy-level diagram of the hydrogen atom is a direct consequence of the inverse-square nature of the electron-proton force. (See Section 35 - 1.) In view of the different energy-level structure, it is clear that the nuclear force must be very different from the Coulomb force.

35–11 THE NUCLEAR FORCE

When we discussed the hydrogen atom, we were able to make the guess that the force between the electron and proton, when they are as close as atomic dimensions, followed exactly the same law as that between charged particles on the macroscopic scale. That is, we assumed that Coulomb's law, which was found by measuring macroscopic forces at macroscopic distances, was still valid on the atomic scale. Rutherford's scattering experiments showed this to be true for alpha particles and nuclei, and the great success of the quantum description of the atoms showed that it was a good guess also for electrons and nuclei. But what about the nuclear force which holds together the deuteron on a still smaller scale? We cannot measure its properties by doing experiments with objects at macroscopic distances, as we could with the Coulomb force. It does not come into play until the neutron is very close to the proton (~ 10^{-14} m). In addition, the nuclear force must be very much stronger than the force between a proton and electron. We can see this by considering the size of the deuteron, which is about 10^{-14} m. We know that for a particle to be contained in a volume of this dimension, its standing wave pattern must have a wavelength of the same order of magnitude. Since $p = h/\lambda$, this means that the neutron and proton must have a kinetic energy of the order of $E_K \approx p^2/M \approx h^2/M\lambda^2 \approx 10^7$ ev or about 10 Mev. Since the binding energy of the deuteron is only 2 Mev, the average potential energy must be of the same order of magnitude as the kinetic energy and must be negative. We can contrast this with the potential energy of a proton in the Coulomb field of another proton at a distance of 10^{-14} m, which is about 0.14 Mev and falls off inversely with distance. As a simple model (only very qualitatively correct) let us say that the potential energy of a neutron and a proton is zero outside some small radius and constant inside. This is the so-called square well potential. Exact calculations with such a potential show that when it is not too strong there is only one possible standing wave pattern corresponding to the ground state of the deuteron.

Just as in the case of the Coulomb potential, we can learn more about the nuclear potential

TABLE 4 *Comparison of the Hydrogen Atom and the Deuteron*

	HYDROGEN ATOM	DEUTERON
Binding energy	13.6 ev	2.2×10^6 ev
Spectrum	Many lines, which suggests many energy levels	Only one line, corresponding to break-up energy
Energy levels	Large number of levels $-13.6/n^2$ ev	Ground level only -2.2×10^6 ev
Size	$\sim 10^{-10}$m	$\sim 10^{-14}$m

by doing the analogue of the Rutherford scattering experiments. That is, we can investigate the scattering of protons on protons and neutrons on protons. Since the discovery of the neutron there has been a great deal of work on this subject, and it has been confirmed that the potential is indeed somewhat like the guess above. To be sure, the potential is more complicated, and in fact physicists are still working on the problem today.

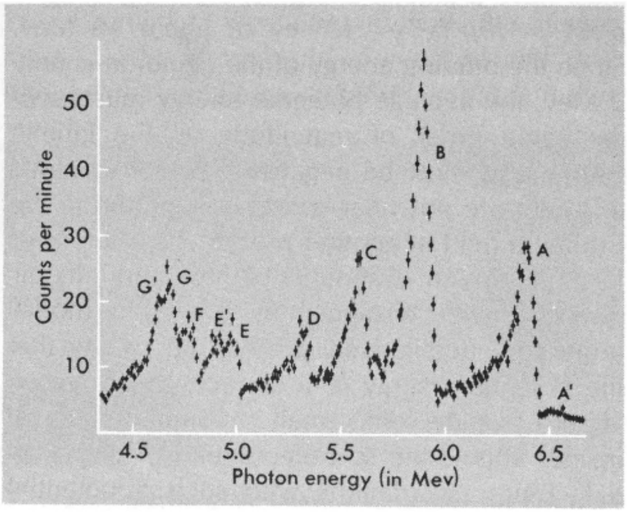
35–12 COMPLEX NUCLEI

Just as there are many different kinds of atoms with different numbers of electrons, there are also many different nuclei with different numbers of neutrons and protons in them. Of course, we know that the number of protons determines the charge of a nucleus and this in turn determines the properties of the atom with the nucleus at its center. The number of neutrons is important in determining the mass of the nucleus and its internal properties. Nuclei with the same proton number and different neutron numbers are called isotopes and can be distinguished by a mass spectrograph (Section 28 – 10).

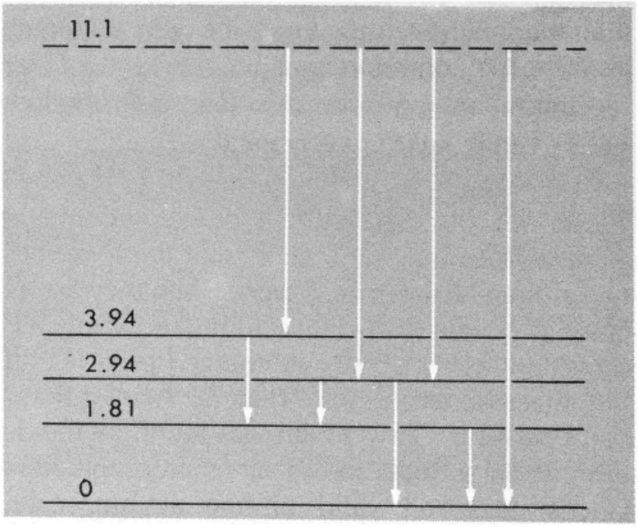
Just as with atoms or molecules, nuclei with many protons and neutrons can be excited to emit photons with characteristic spectra. Unlike the deuteron, the spectra of complex nuclei contain many lines, corresponding to energies in the range of 0.1 – 10 Mev.

One of the most direct ways of exciting the spectrum of some nuclei is to use the method of neutron capture. Virtually every nucleus existing in nature is capable of capturing slow neutrons to produce the next isotope with one more neutron. In doing so it emits its characteristic spectrum of "neutron capture" gamma rays. As the nuclei get heavier, these spectra become more and more complicated, and one of the challenging problems in this field is to design gamma-ray spectrographs which are sufficiently sensitive and have good enough resolution to see all of the lines. Figure 35 – 16 shows the photon spectrum of a mercury nucleus. The very fact that there are discrete lines in the spectra shows us that in discussing nuclear physics we must use quantum ideas, in particular the idea of discrete energy levels. Figure 35 – 17 shows the energy-level diagram of one of the isotopes of magnesium. Notice that five energy levels account for eight observed lines.

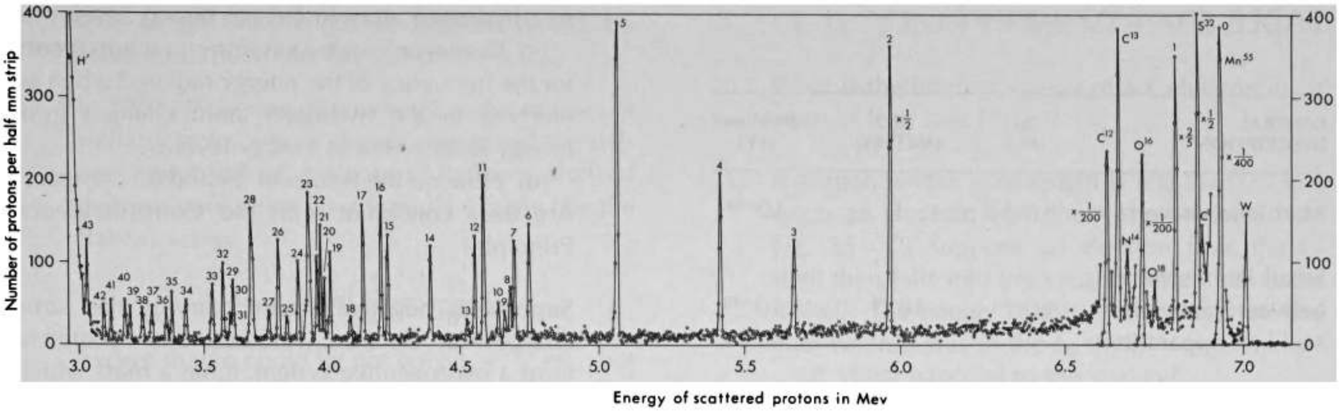
The emission spectra are not the only tools available to us for studying the energy levels of nuclei. We can also do scattering of the type which occurs in the Franck-Hertz experiment. In the nuclear case, we scatter energetic protons from various nuclei and investigate the inelastic collisions. If the protons have kinetic energy below that necessary to produce the first excited state of the target, they will be scattered elasti-



35–16 Part of an actual gamma-ray spectrum produced when neutrons are captured on mercury. The peaks correspond to the various photon lines. (From: Kinsey and Bartholomew, Canadian Journal of Physics, 31, 1025, 1953.)

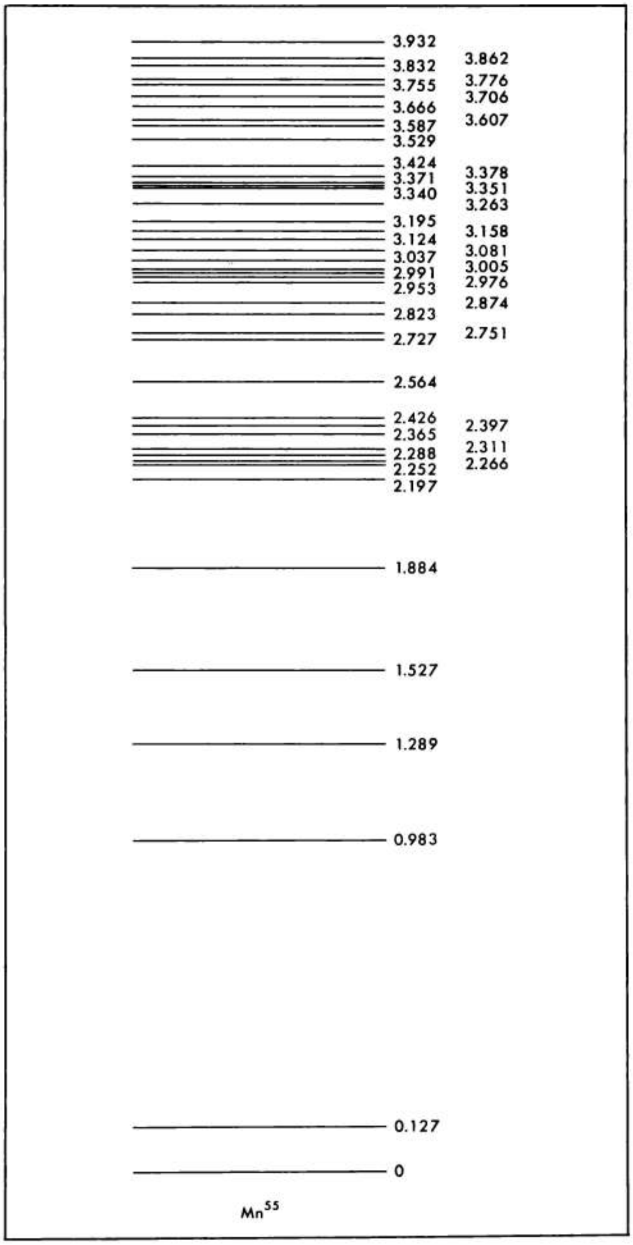


35–17 Part of the energy-level diagram of magnesium (Mg^{26}) showing some of the photon lines seen in the slow neutron capture reaction. The dotted line corresponds to the binding energy of the neutron. The energies are in Mev.



35–18 The spectrum of scattered protons produced when a beam of 7.03 Mev protons from an electrostatic accelerator is incident on a thin target of manganese. The protons were scattered at an angle of 50° to the incident beam and were then deflected magnetically. Thus protons of different energy hit different locations on the photographic film. The points on the graph represent the number of protons landing at various places on the film. These are obtained by looking at the film under a microscope and counting the number of proton tracks as a function of distance along the plate. Most of the peaks correspond to the different excited states of manganese. Note that 43 energy levels of manganese have been seen in this experiment. Some of the peaks correspond to elastic scattering from various other elements which contaminate the target or exist in its backing material. Using conservation of energy and momentum for elastic scattering from nuclei with various masses, we can compute exactly where these contaminant peaks should be. They are labeled in the figure. (From: Mazari, Sperduto, and Buechner, Physical Review, 108, 103, 1957.)

cally. As soon as the kinetic energy is above this threshold, inelastic scattering will occur, and some of the protons will come out with a definite loss of kinetic energy. In the case of electrons in the Franck-Hertz experiment, this was measured by a change in the current of electrons. For the case of protons on nuclei the experiment is far more sensitive, because it is quite possible to measure directly the energies of the scattered protons at any given angle. In Fig. 35 – 18 we show the energy spectrum of 7 Mev protons scattered from manganese. There are discrete lines in the energy spectrum of the scattered protons. That is to say that the protons cannot lose arbitrary amounts of energy on colliding with a nucleus, but only those amounts corresponding to the actual energy levels of the target nucleus. In fact, we can read the possible excitation energies of the target nucleus right off the inelastic proton spectrum. The energy levels are shown in Fig. 35 – 19.



35–19 The energy-level diagram of the nucleus Mn^{55} as obtained from inelastic proton scattering. The energies are given in Mev.

TABLE 5

GENERAL DESCRIPTION	m (KG)	d (METERS)	$E \approx h^2/md^2$ (EV)
Marble in drawer	10^{-2}	1	10^{-46}
Small latex spheres between plates in Millikan apparatus	10^{-6}	10^{-2}	10^{-38}
A hydrogen atom in a molecule	10^{-27}	10^{-10}	10^{-1}
An electron in an atom	10^{-30}	10^{-10}	10^{2}
A neutron in a nucleus	10^{-27}	10^{-14}	10^{7}

Comparing this figure with Fig. 33 - 5, which shows the energy levels of the mercury atom, we again notice the factor of 10^6 in the energy scale. One thing seems to be quite clear. Nuclei are quantum systems of neutrons and protons, and the basic reason for the high energy scale is the small size of the nuclei. The connection between the size of a system and the spacing of its energy levels is shown in Table 5.

Even for systems which can barely be seen under the microscope, the energy levels are so closely spaced that they cannot be resolved. This is why man never noticed them before he started to probe into the world of the atom.

QUESTIONS

1.* The calculation of energy levels of a hydrogenlike atom assumed a nonrelativistic form for the kinetic energy. Justify the validity of this assumption, using your knowledge of the rest energy of the electron and the Rydberg.

2. (a) Derive an expression in terms of fundamental constants for the velocity v_n of an electron in the nth energy level of a hydrogen atom. (See Section 34 - 11.)

(b) Classical theory predicts that an electron moving in a circle should radiate, the frequency of the radiation being equal to the orbiting frequency. Derive an expression for this frequency in a hydrogen atom in the nth energy level.

(c) Derive an expression from quantum theory for the frequency of the energy radiated when an electron in the hydrogen atom changes from energy level $n + 1$ to energy level n.

(d) How do the results of (b) and (c) compare? Are they consistent with the Correspondence Principle?

3. Suppose a negative muon, carrying the same charge as an electron, were bound to a proton to form a hydrogenlike system. It has a mass which is 212 times the electron mass. What would you predict about

(a) the binding energy of the system?

(b) the wavelength of the photon emitted when the muon changes energy from $n = 2$ to the ground state? (Spectra from muonic atoms have been observed even though they exist for only microseconds or less.)

4. The very short-lived "atom" positronium consists of an electron and a positron bound together. What do you predict about

(a) the binding energy of the system?

(b) the wavelength of the photon emitted when positronium changes energy from $n = 2$ to the ground state?

5. Calculate the change in angular momentum with respect to the nucleus for the transition of an electron in a hydrogenlike orbit from the level $(n + 1)$ to the level n. Does it depend on n? Does it depend on Z?

6.* The first three spectral lines of the Lyman series of hydrogen are 1216Å, 1026Å, and 973Å. To which helium ion lines in Fig. 35 - 2 do these correspond?

7.* What is the approximate radius of the helium ion in its ground state?

8.* In what region of the spectrum—infrared, visible, or ultraviolet—would you expect to find lines of neutral helium whose wavelengths are close to lines in the hydrogen spectrum?

9. What is the smallest value of n for which $(E_n - E_{n-1})$ of the helium ion will produce a photon in the visible part of the spectrum?

10. Suppose that you have singly ionized helium with the ion in its ground state.

(a) How much energy in ev is required to strip away the remaining electron?

(b) What wavelength of photon would be needed to knock out this electron?

11. In 1897, Pickering discovered a series of lines in stellar spectra which closely resembled the Balmer spectrum of hydrogen. Rydberg showed that this new series fits a formula exactly like the Balmer series,

$$E_{\text{photon}} = R\left(\frac{1}{2^2} - \frac{1}{m^2}\right)$$

except that m could be not only 3, 4, 5, etc., but also $2\frac{1}{2}$, $3\frac{1}{2}$, $4\frac{1}{2}$, etc. Bohr finally showed in 1913 that these were not lines of hydrogen but were due rather to the helium ion. Reconstruct his argument. (Hint: try transitions from the level n to the level 4.)

12. The spectrum of atomic helium contains a line of wavelength $\lambda = 584$ Å. If the lower energy level in the transition is the ground state, what is the upper level and what is Z_{eff} for it?

Mathematically, there is more than one answer to this question. However, only one answer is physically acceptable. Which one, and why?

13.* Could any spectral line in Table 1 correspond to a transition in doubly ionized lithium to the $n = 2$ state from a higher energy state?

14. Does the energy in the last row in Table 1 correspond to a transition to the ground state of doubly ionized lithium? What is the initial state for this transition?

15. Identify the levels involved in the transitions yielding the lines of the lithium spectrum shown in Table 2.

16. The spectrum of normal lithium has a fairly intense line in the ultraviolet at 2741Å. If the photon comes from a transition to the ground state, what was the excited level?

17. You have seen that the Z_{eff} for the first electron removed from helium is 1.35. Also, the Z_{eff} for the second electron removed from lithium is 2.36. Extrapolate these data to predict the third ionization energy of beryllium, which has a nuclear charge of four elementary charges. Explain your reasoning.

18. What kind of an experiment would you perform to measure the charge on the nucleus of a lithium atom?

19.* Why do we not expect to find any data points in Fig. 35 – 8 above the line $\sqrt{E} = \sqrt{R}\,Z$?

20.* What is the binding energy of a K electron in an atom of lead (see Fig. 6)?

21. A photon whose wavelength is less than 0.141Å ejects an electron from the K-shell in lead. (See Fig. 35 – 7.) Suppose an electron from the L_{II} shell then falls into the vacant "hole" left in the K-shell. The energy of this electron will be less than when it was in the L_{II} shell.

(a) What happens to this energy?

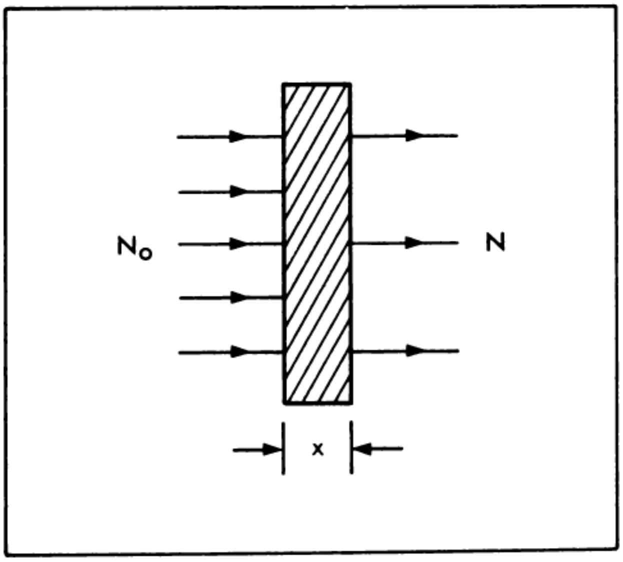
(b) How much is it?

22. With element of atomic number 12 (magnesium) an X-ray absorption edge occurs at $\lambda = 197$Å, for an electron with $n = 2$.

(a) Calculate Z_{eff}.

(b) Explain why this value is different from that shown in Fig. 35 – 8.

23. The absorption coefficient, μ, whose values are plotted in Fig. 35 – 7, is defined in the following way: Consider a beam of X rays striking a sheet of material whose density is ρ and whose thickness is x. The number of photons striking the target is N_0 and the number passing directly through without being absorbed is N (Fig. 35 – 20). Experiments done with different thicknesses x show that a monochromatic beam of X rays is absorbed according to the following law: $N/N_0 = e^{-\mu\rho x}$. What is the ratio of the fraction of photons absorbed from an X-ray beam of $\lambda = 0.4$ Å incident on a 1.0×10^{-2} cm thick sheet of lead to the fraction absorbed from a beam of $\lambda = 0.6$Å incident on the same sheet?



35–20 For Problem 23.

24. We can assign an atomic number to an element on the basis of its serial number in the chemical list in which elements are arranged in the order of increasing atomic masses. If these masses are determined by chemical measurements obtained from natural sources, nickel would have an atomic number of 27 and cobalt an atomic number of 28. However, if we use the X-ray absorption edges to determine the atomic number, the numbers are 28 for nickel and 27 for cobalt. How can you account for this apparent contradiction? (Hint: the natural isotopes of nickel have atomic masses of 58, 60, 61, 62, and 64 amu, while cobalt has only one natural isotope, whose mass is 59 amu.)

25. When high-energy photons between 100 and 200 Mev are incident on hydrogen there occurs a phenomenon similar to the X-ray absorption edge in atoms. Below 150 Mev photons are scattered. At 150 Mev absorption begins. It has been shown that this is due to the creation of a new particle called the pion. The fact that pions are not produced by photons with less energy is because the pion has a rest energy. What is the rest energy of the pion? (If the pion is neutral, the proton is unchanged; if the pion is positively charged, the proton becomes a neutron with substantially the same mass.)

26.* To compute Z_{eff} from Fig. 35 – 9 for the elements between Na and Ar, what value of n would you use?

27.* Referring to Fig. 35 – 9, what is the value n for the outermost electrons of krypton?

28. Using data from Fig. 35 – 9, calculate the effective nuclear charge acting on the outermost electron (a) in an argon atom and (b) in a potassium atom.

29. Calculate the radii of krypton and rubidium atoms.

30.* If Fig. 35 – 7 were extended to include longer wavelengths, would you expect to see more absorption edges?

31. In a sample of helium gas containing many helium atoms, all the electrons are described by a wave pattern with $n = 1$. Does this violate the Pauli principle? Explain.

32.* One electron in a box has a standing wave pattern described by the quantum numbers (3, 1, 1). Another electron in the same box has quantum numbers (2, 2, 2). How do their energies compare?

33.* How many electrons could exist in the same box with the same energy as the (3, 1, 1) electron in the preceding problem? How many with the same energy as the (2, 2, 2) electron?

34.* Measure the ratio of the distances along the plate in Fig. 35 – 15. From this obtain the energy of the neutron-proton capture gamma ray, using the known energy of the positron annihilation photons (0.511 Mev).

35. In the helium nucleus (alpha particle) there are two protons and two neutrons. The particles are separated on the average by a distance of 10^{-14} m.

(a) What is the average electrical force between the protons?

(b) Since the electric force is a repulsive force, there must be an attractive force in the nucleus to hold the protons together. Can it be the gravitational force? Can it have the same dependence on the distance as the electric force?

36. Nuclear radii are of the order of 10^{-14} m. Suppose an electron is confined in a nucleus, then its momentum would be about the same as that of a neutron in a nucleus.

(a) Estimate the kinetic energy of such an electron. (Caution: Which relation between momentum and kinetic energy should you use, the nonrelativistic or the relativistic?)

(b) How large a potential would it take to keep the electron inside the nucleus?

(c) Do you think it is reasonable that an electron could be part of a nucleus?

37.* The peaks in Fig. 35 – 16 correspond to differences in the nuclear energy levels of mercury. The peaks in Fig. 35 – 18 correspond directly to the energy levels of manganese. Why?

38.* The peak at 6.94 Mev in Fig. 35 – 18 is associated with elastically scattered protons from the ground state of Mn^{55}. But the incident proton energy was 7.03 Mev. Where did the 0.09 Mev go?

39. Sketch a graph of the number of scattered protons as a function of energy, like Fig. 35 – 18 if incident protons of 2.00 Mev energy were scattered by Mn^{55}.

40. Figure 35 – 18 gives the number of scattered protons as a function of their energy when 7 Mev protons are incident on a manganese target. Some of the peaks are due to elastic scattering from impurities in the target. Why are the peaks from the lighter impurities always to the left of those from the heavier impurities?

41. Figure 35 – 19 shows the energy levels of the manganese nucleus. What are the possible energies of the photons which may be emitted when the nucleus is in its fourth excited state?

42. Free neutrons do not last very long in the presence of matter. They are captured by the nuclei, emitting photons. For example, when neutrons are captured in Ca^{40}, Ca^{41} is formed and photons with energies of 6.4, 5.9, 1.9, and 0.5 Mev are observed. It is also known that the binding energy of the last neutron in Ca^{41} is 8.3 Mev. Construct (the best you can) an energy-level diagram for Ca^{41}.

36

Change in Atoms and Nuclei

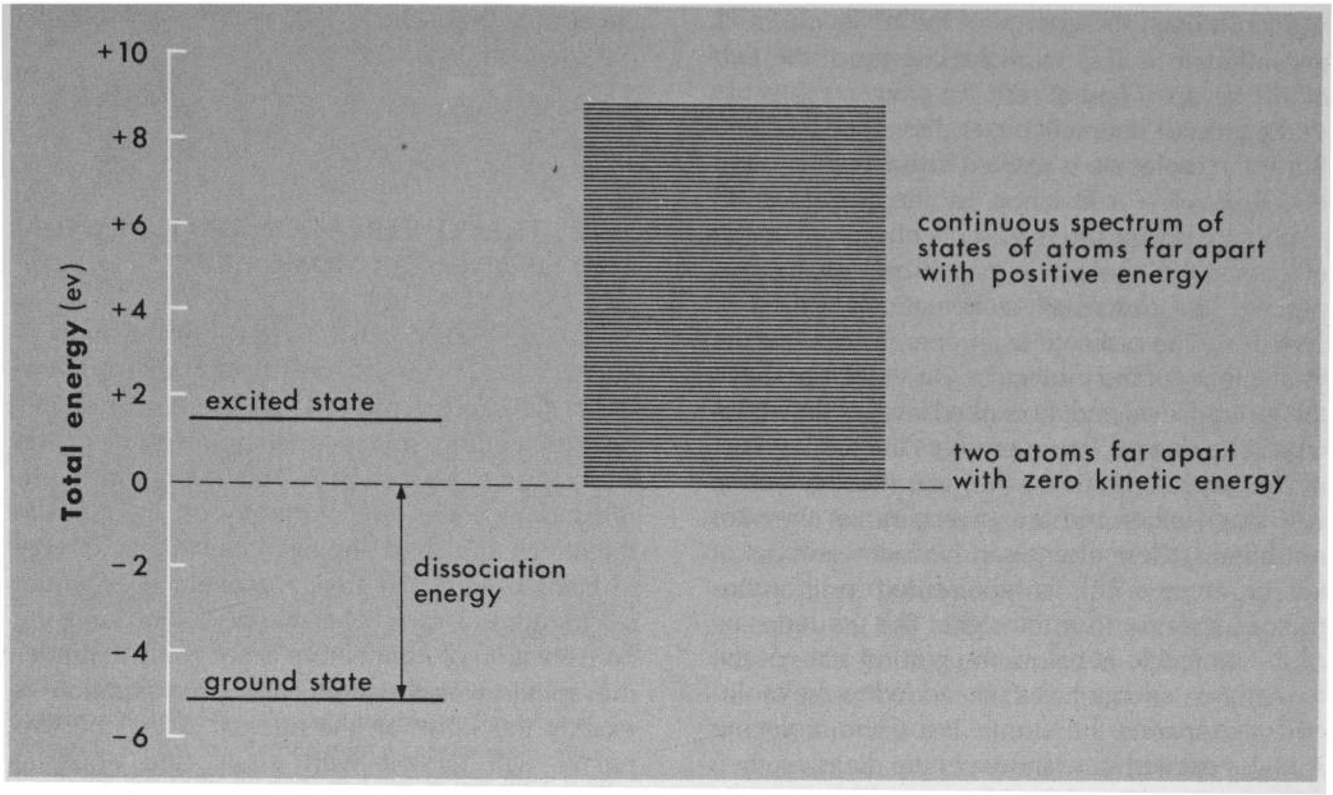
WE HAVE emphasized until now the stability of atoms, the sameness of the energy levels in atom after atom of the same element, the definite unvarying energies of the levels of atom, molecule, and nucleus. But we know very well that the world does not stand still, especially the world of the atom and the nucleus. Indeed, what we know about their energy levels is learned mainly through observing atoms changing from one level to another, very often emitting a photon. We now have to look a little more closely than we have yet done at how changes can take place in quantum systems.

Why do some nuclei decay through radioactivity? Why is that decay rate so independent of external conditions and always exponential in time? Why do some molecules combine and others decompose? How, when, and why does an atom come to leave an excited state by the emission of a photon? All of these questions are central to the understanding of the world, for change is no less real than stability.

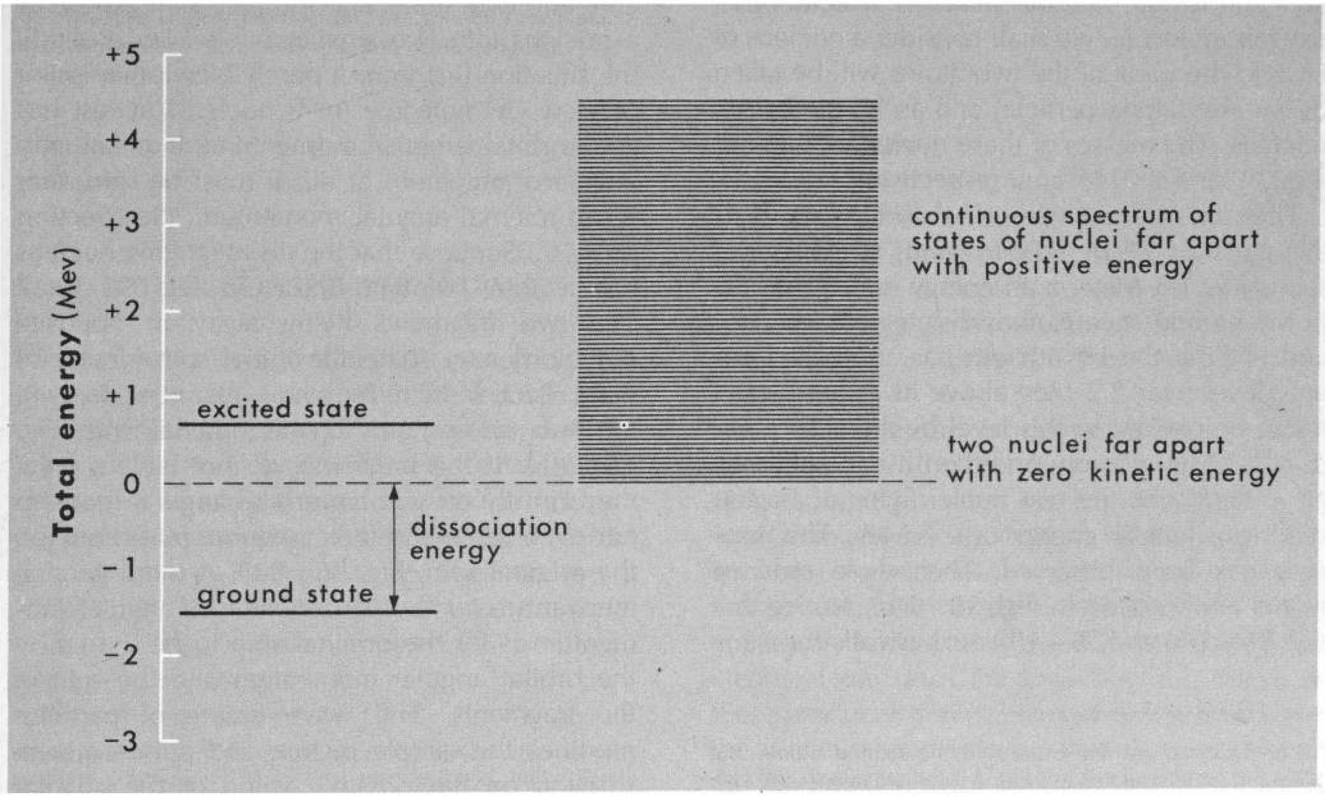
We shall try to approach these questions, which are difficult ones, by simple and approximate means. We shall argue quantum mechanical questions by drawing upon our imperfect but very helpful wave analogies —the waves in a string and in water—to discuss what happens to probability waves, which is the subject of our interest in this chapter. We shall also see how crucial and how simple are the results of the conservation laws, which cannot by themselves tell us which reactions are fast and which slow, but which do lay down the ground rules that determine which reactions may or may not take place.

36–1 THE CONSERVATION OF ENERGY

If a system can possibly split into two parts, which then fly apart with kinetic energy such that the total energy remains unchanged, sooner or later it will do so of its own accord. We shall illustrate this with two examples in vastly different ranges of energy. First, consider an oxygen molecule, O_2. Since energy must be supplied to separate the two


+10
+8
+6
+4
+2
0
-2
-4
-6
Total energy (ev)
excited state
ground state
dissociation energy
continuous spectrum of states of atoms far apart with positive energy
two atoms far apart with zero kinetic energy

36–1(a) The schematic energy-level diagram for the breakup of an oxygen molecule into two atoms.


+5
+4
+3
+2
+1
0
-1
-2
-3
Total energy (Mev)
excited state
ground state
dissociation energy
continuous spectrum of states of nuclei far apart with positive energy
two nuclei far apart with zero kinetic energy

36–1(b) The schematic energy-level diagram for the breakup of a Li^6 nucleus into an alpha particle and a deuteron.

oxygen atoms, the energy of the molecule in its ground state is less than the energy of the two atoms far apart and at rest. An oxygen molecule in the ground state will never dissociate by itself. But if the molecule is excited into a high-enough energy level—for instance, by the absorption of a photon or by an inelastic collision with an electron—the molecule may break apart. This process is described schematically in Fig. 36 – 1(a). The ordinate is an energy scale for the total energy of the molecule. On the left we have the ground level and an excited level of the molecule. The dashed line represents the total energy of the two atoms when they are not interacting with each other and are at rest. When they are not bound, these atoms can have any amount of kinetic energy. This is represented by the continuous spectrum on the right. The ground state of the molecule is below the ground state of the two atoms; energy has to be added to the molecule to separate the atoms. But if the molecule is in the excited state shown in the diagram, then there is enough energy to separate the two atoms and give them some kinetic energy.

Similar processes occur in the world of the nucleus. We can use a diagram like that in Fig. 36 – 1(a) to illustrate the situation. Instead of an oxygen molecule, we shall consider a nucleus of Li^6 and the parts of the two atoms will be taken by an He^4 (alpha particle) and an H^2 (deuteron) nucleus. The masses of these nuclei are 6.0151, 4.0026, and 2.0141 amu respectively.*

The sum of the masses of H^2 and He^4 exceeds the mass of Li^6 by 0.0016 amu, which corresponds to 1.5 Mev on an energy scale. Thus Li^6 in the ground state cannot disintegrate into He^4 and H^2. But the Li^6 nucleus has an excited energy level near 2.2 Mev above its ground state. It can be excited to this level by the absorption of a 2.2 Mev photon, and then it can split into $H^2 + He^4$, with the two nuclei flying apart and sharing a kinetic energy of 0.7 Mev. This process has been observed. The whole state of affairs is illustrated in Fig. 36 – 1(b). Notice that Fig. 36 – 1(a) and 36 – 1(b) are basically the same except for the factor of 10^6 on the energy scale. (See Section 35 – 10.)

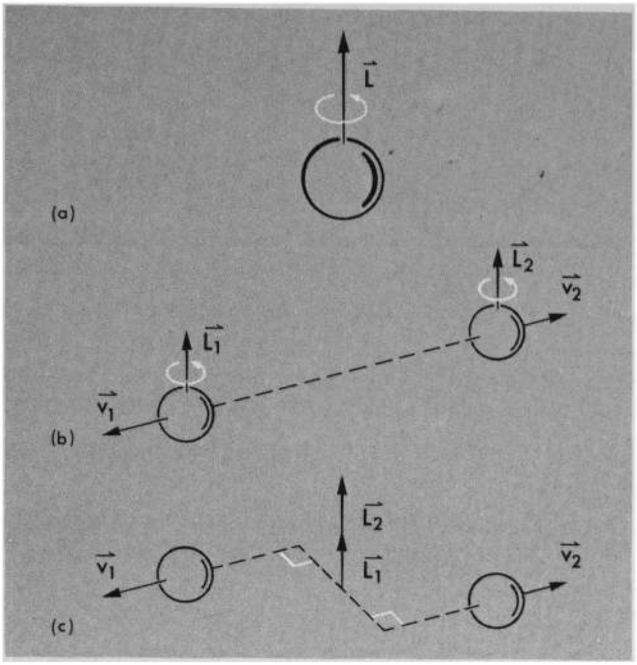
36–2 THE CONSERVATION OF MOMENTUM AND OF ANGULAR MOMENTUM

So far we considered only the conservation of energy. What about conservation of momentum? An excited oxygen molecule at rest has no momentum, so the total momentum of the atoms must also be zero. This condition is fulfilled only if the two atoms fly off in opposite directions and share the available kinetic energy in such a way that their momenta have equal magnitudes. This is what happens, and thus the conservation of momentum is not going to hinder the splitting of the molecule. The situation is exactly the same on the nuclear scale; the two nuclei will fly off with equal and opposite momenta. This can be checked experimentally.

Like the conservation of linear momentum, the conservation of angular momentum can also be satisfied, both for the dissociation of a molecule into two atoms and for the disintegration of a nucleus into two smaller nuclei. Let us study the situation first from a purely Newtonian point of view. A molecule (or a nucleus) at rest has zero orbital angular momentum. If it has any angular momentum at all, it must be spin, that is, an internal angular momentum. (See Section 19 – 10.) Suppose that the disintegrating nucleus has a spin pointing up as in Fig. 36 – 2(a). The two fragments flying away in opposite directions may share the initial spin in several ways. Each of them may have a spin of its own, the two adding up to the original spin Fig. 36 – 2(b). If the fragments do not have a spin, they can fly off with an orbital angular momentum with respect to their center of mass equal to the original spin, Fig. 36 – 2(c). A more general mechanism for the conservation of angular momentum is for the original spin to be shared by the orbital angular momentum and the spin of the fragments. The wave nature of particles modifies this simple picture and puts some restriction on the possible value of the angular momentum, but the mechanisms for conserving

---

*Actually these are the masses of the neutral atoms. But since the number of electrons in Li is the same as that of H + He, we need not be concerned with them. See Chapter 32, Table 2.



36–2 The conservation of angular momentum when a system with spin breaks up into two parts. The original system in (a) has spin but no orbital angular momentum. In (b) the fragments fly apart in a line so that there is no orbital angular momentum. Each fragment has spin, and their total spin adds up to the original spin. In (c) the fragments fly apart along two parallel lines so that they have an orbital angular momentum. Since neither of them has spin, this orbital angular momentum equals the original spin angular momentum. The general case is for the fragments to have both spin and orbital angular momentum.

angular momentum discussed above remain possible.

36–3 OTHER CONSERVATION LAWS

In all observations to date, electric charge has always been conserved. But the conservation of electric charge, energy, momentum, and angular momentum are not the only factors which determine whether a reaction can occur or not. For example, compare a ${}_{36}\mathrm{Kr}^{78}$ nucleus* to the pair of nuclei ${}_{20}\mathrm{Ca}^{45}$ and ${}_{16}\mathrm{S}^{32}$. The masses are 77.945, 44.967, and 31.981 amu respectively. The sum of the last two masses is much less than the first; energetically the ${}_{36}\mathrm{Kr}^{78}$ nucleus can disintegrate into ${}_{20}\mathrm{Ca}^{45}$ and ${}_{16}\mathrm{S}^{32}$. In fact, the available energy is more than 10^3 Mev. Energy, momentum, angular momentum, and electric charge can easily be conserved in the reaction. We may then ask what prevents the ${}_{36}\mathrm{Kr}^{78}$ nucleus from disintegrating into ${}_{20}\mathrm{Ca}^{45}$ and ${}_{16}\mathrm{S}^{32}$. Since the number of neutrons in ${}_{20}\mathrm{Ca}^{45} + {}_{16}\mathrm{S}^{32}$ is one less than in ${}_{36}\mathrm{Kr}^{78}$, one neutron would disappear in this reaction. Its rest energy of 10^3 Mev would then be shared by the two fragments in the form of kinetic energy, or perhaps a high-energy photon would be emitted. In both cases, it would be quite easy to observe this reaction. It has never been observed. Nor has any other splitting been seen in which the energy released is equal to the rest energy of a neutron. Nor have reactions been observed in which the number of protons is reduced and their rest energy released in the reaction. All our experience to date indicates that the total number of neutrons plus protons, or nucleons for short, has to be conserved. This is another fundamental conservation law which must be added to the conservation of energy, momentum, angular momentum, and electric charge.

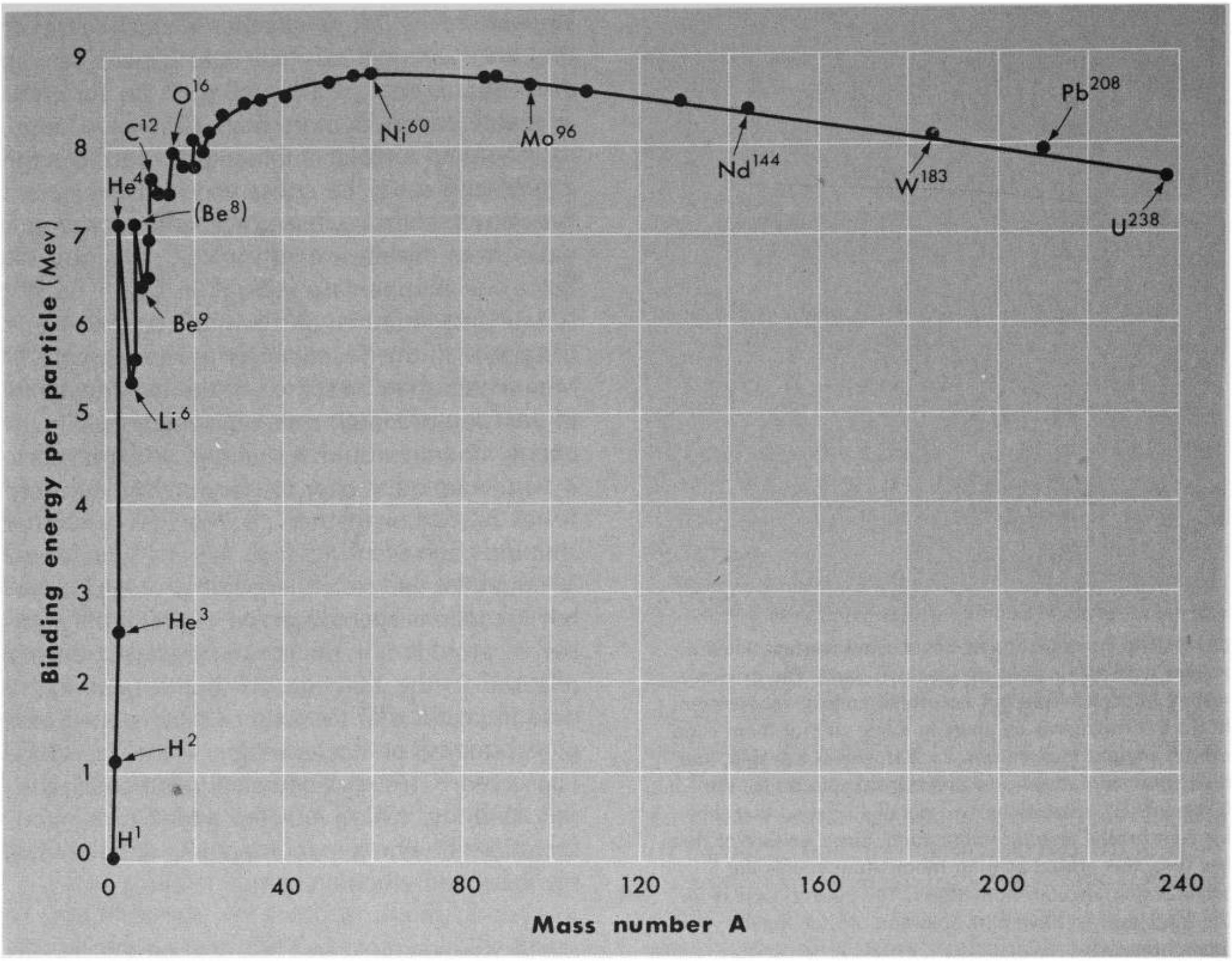
36–4 FISSION

The conservation of energy prevents a light nucleus like Li^6 in its ground state from disintegrating. If the nucleus is sufficiently excited, it splits easily, as we have seen in Section 36 – 1. The situation is quite different with the very heavy nuclei.

For example, the uranium nucleus with mass number $A = 236$ might be broken up and rearranged into two other nuclei, one of $A = 90$ (which might be an isotope of the metal zirconium), and one of $A = 146$ (which might be the rare element neodymium). Since $90 + 146 = 236$, we have exactly the same number of nucleons. How does the total binding energy of these products, that is, the energy needed to take the nucleus apart into nucleons, compare with the binding energy of uranium 236 in its ground state? The binding energies per nucleon can be

---

*The symbol ${}_{36}\mathrm{Kr}^{78}$ is a compact notation for the krypton nucleus containing 78 protons and neutrons, that is, with mass number $A = 78$. The nuclear charge for all krypton isotopes is $Z = 36$.



36–3 The average binding energy per particle for the naturally occuring nuclei plotted as a function of mass number A. Be^8 has been included, even though it does not occur naturally, because of our interest in it in this chapter.

roughly read from the curve in Fig. 36 – 3, but we give them here more accurately to two significant figures.

		Total binding energies in Mev
U^{236}	236 nucleons at 7.5 Mev/nucleon	1800
Zr^{90}	90 nucleons at 8.7 Mev/nucleon	780
Nd^{146}	146 nucleons at 8.2 Mev/nucleon	1200

On adding up the binding energy in the two smaller nuclei, we get 1980 Mev. Consequently, in rearranging the nucleons we get a surplus of binding energy of about 200 Mev. Since the binding energy is now larger, the nuclei have given up 200 Mev. This 200 Mev must have been transferred into energy in some other form. Thus, if any way at all exists for a U^{236} nucleus to rearrange itself into a pair of separated parts, one a Zr nucleus and the other a Nd nucleus, the two fragments will fly apart with a kinetic energy of about 200 Mev. Actually, some of the energy is always used up in exciting the fragments or in freeing a few odd neutrons or an alpha particle, but most of it does turn up as kinetic energy of two big fragments.

This energetic act of splitting, called fission, takes place in heavy nuclei. Roughly speaking, the nucleus divides into two large fragments, a little like an amoeba. The change goes on spontaneously, according to an exponential law of decay. That is, the number of decays per unit time is proportional to the number of undecayed nuclei. For uranium nuclei in their ground states, the rate is almost unbelievably slow.

The decay takes a time which is a billionfold the whole span of the earth's history. Yet there are excited states of these same uranium nuclei which split apart in 10^{-16} second or less, releasing a few neutrons in the process. It is these states which make possible the chain reaction which releases nuclear energy on a large scale in a nuclear bomb or in a nuclear reactor.

The binding-energy curve of Fig. 36 – 3 suggests that not only one particular kind of fission but a wide variety of splittings might be expected. Indeed, even such an unequal splitting as the chipping off of an alpha particle, $A = 4$, is possible in many nuclei. Every nucleus of high A can decay of its own accord into a nucleus with four fewer nucleons, and emit an alpha particle. All of these nuclei are said to be naturally radioactive. In fact, most of the nuclei above $A \approx 140$ are unstable, but the half-lives of those lighter than lead are so great that the activity cannot usually be observed (even though there are ways to observe radioactive emission of alpha particles from nuclei decaying so slowly that their half-lives are around 10^{14} years). Actually, there are only a few exceptional alpha-decaying nuclei lighter than lead whose alpha decay has been observed.

Obviously, the conservation laws really do not tell the whole story of decaying systems. To see what kind of additional factors are involved and how they come into play, we shall concentrate on one process—alpha decay.

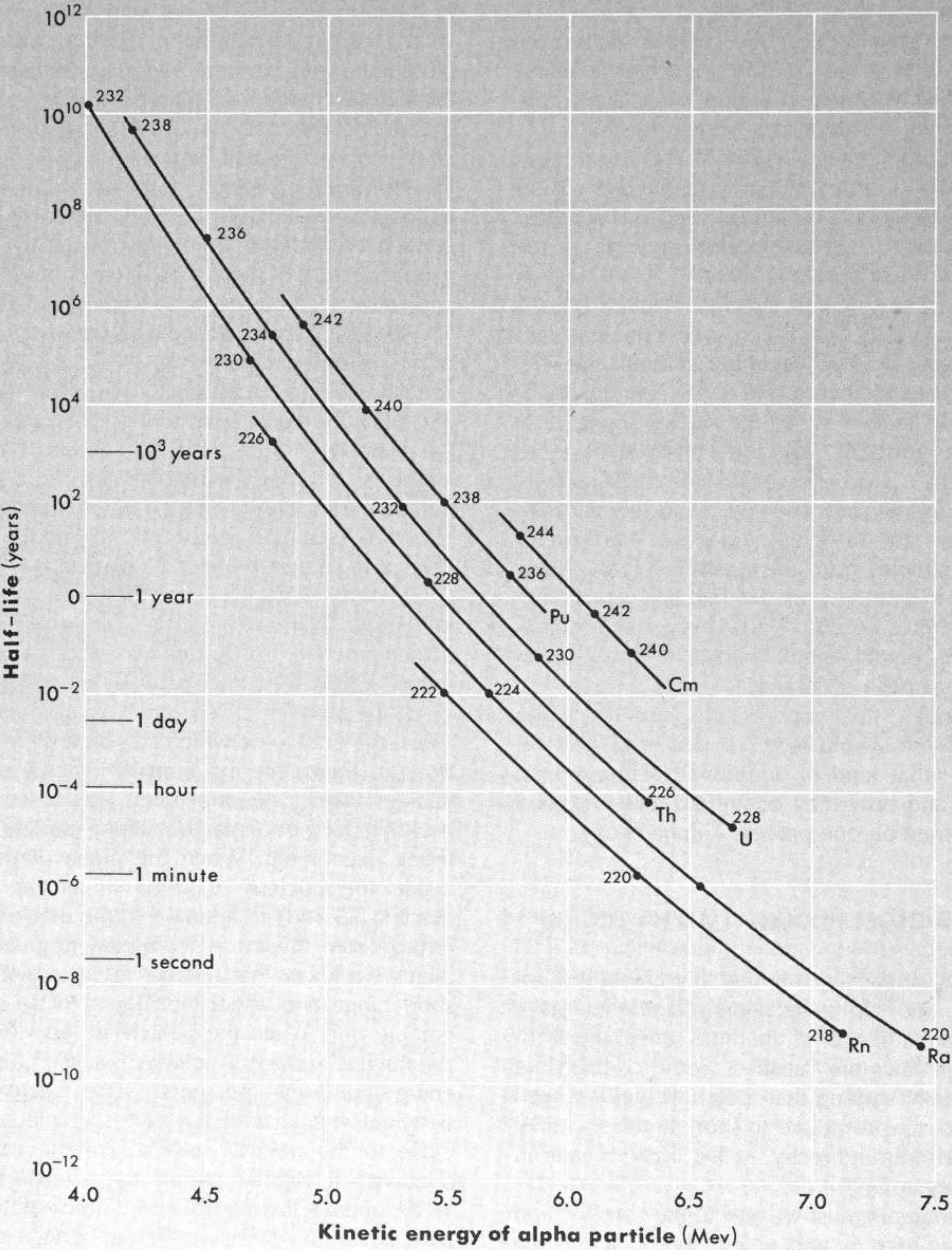
36–5 A CLOSER LOOK AT ALPHA DECAY

We have already mentioned that among the fissioning and alpha-decaying nuclei there is a tremendous range of lifetimes once the conservation laws are satisfied. Some of them are presumably so long that they have not yet been detected, and some are so short that they can be deduced only indirectly, as we shall see later in this chapter.

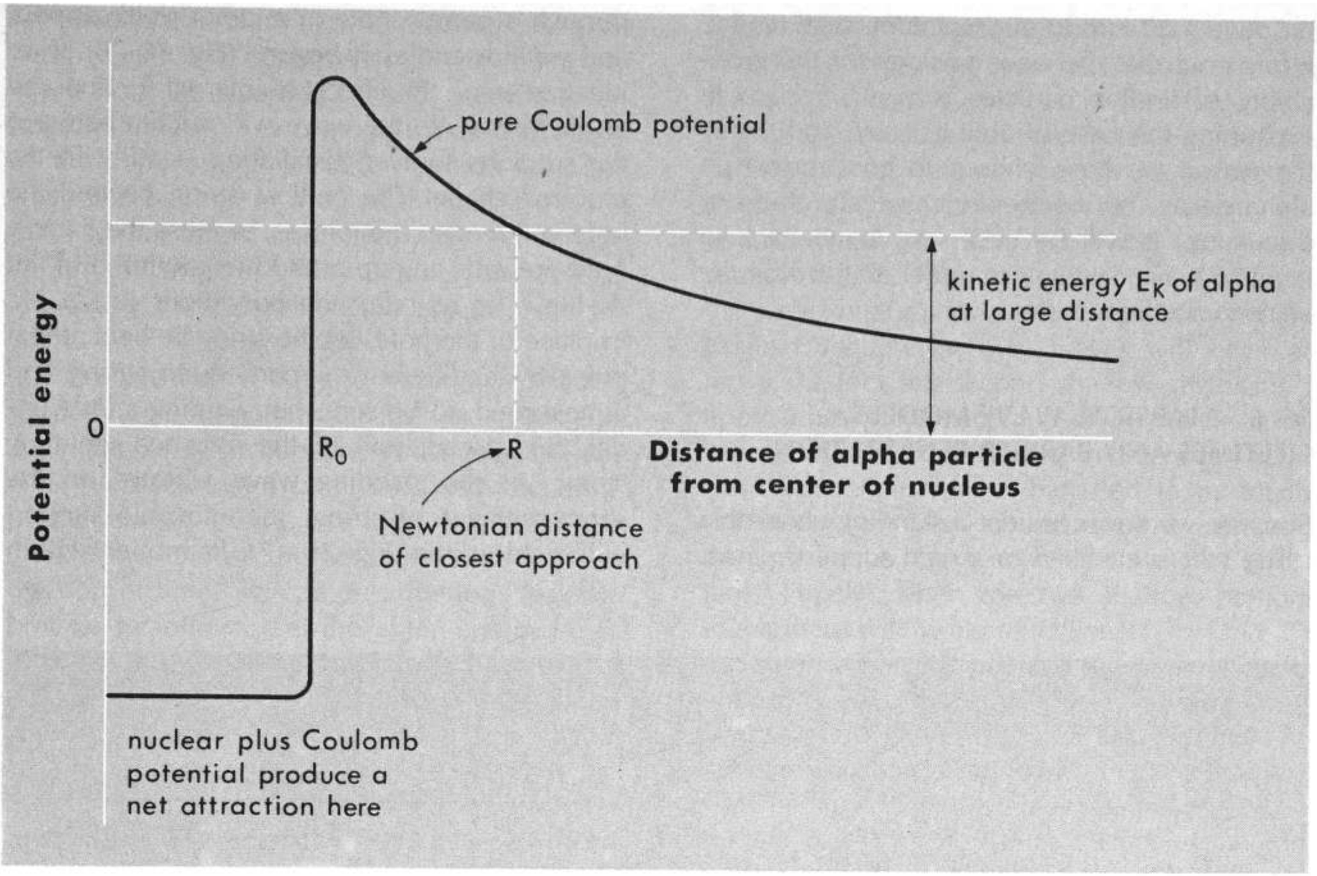
What factors must we add to the conservation laws in order to understand these facts? One additional factor which is always very striking in somehow determining the lifetimes of decaying systems is very well illustrated in the case of alpha decay, where a large body of data has been accumulated. In Fig. 36 – 4 we have plots of the logarithm of the half-life of various alpha emitters as a function of the kinetic energy of the alpha particles. The different plots correspond to different sets of isotopes with the same nuclear charge (radon, radium, thorium, uranium, plutonium, and curium). We can see here that for a fixed charge the alpha-particle kinetic energy alone seems to control the rate of decay. That is to say, the half-life, when plotted against the disintegration energy, forms a smooth curve as long as the charge is fixed. The strong dependence of half-life on energy release is a very general feature of decays in all types of systems, and so we shall pursue our example of alpha decay to see if we can get some sort of explanation of this effect.

In the case of alpha decay we also note another feature. For a fixed kinetic energy release the nuclear charge controls the rate. For example, at 6.3 Mev the half-life goes from about a minute for radon to about a month for curium. This is a factor of nearly 10^5 in lifetime for a change in charge of about 10 percent.

How can we understand these effects? First, let us look a little more closely at the dynamics of our problem. To do this we ask what sort of forces act on the alpha particle. We know that when the alpha particle is far away from the nucleus, the only force which acts on it is the electrostatic force, coming from the charge of the nucleus. This, you remember, was discovered by Rutherford in his famous alpha-particle scattering experiment. When the alpha particle is inside the nucleus, there is a nuclear force (Section 35 – 11) in addition to the electrostatic force. From heavy particle-scattering experiments we know that nuclear forces are of very short range and affect the motion of an alpha particle only when the particle is very close to the nuclear surface. It follows that the potential energy curve has the shape of the electrostatic potential energy until the alpha particle is very close to the nuclear surface. For the case of a heavy nucleus like curium the radius is about 10^{-14} m and the repulsive Coulomb potential rises to about 15 Mev near the surface of the nucleus. On the other hand, an ejected alpha particle, when it is very far from the curium nucleus where the potential energy is negligible, has 6.3 Mev of kinetic enegy. If the alpha particle is to exist for any length of time inside the nucleus



36–4 The half-life of some alpha emitters as a function of the kinetic energy of the alpha particles. The different curves correspond to different families of isotopes.


pure Coulomb potential
kinetic energy E_K of alpha at large distance
Potential energy
0
R_0
R
Distance of alpha particle from center of nucleus
Newtonian distance of closest approach
nuclear plus Coulomb potential produce a net attraction here

36–5 A schematic representation of the potential energy of an alpha particle and a heavy nucleus. Outside the nuclear radius R_0 the potential is that due to the repulsive Coulomb force. Inside R_0 the attractive nuclear force comes into play and the net potential energy becomes negative. For ordinary alpha emitters the kinetic energy at infinity E_K is smaller than the Coulomb potential energy at the nuclear radius R_0 and is shown by the horizontal dashed line in the figure. According to Newtonian mechanics, for any E_K there is a corresponding distance of closest approach R, which is also indicated on the figure.

with a positive total energy of 6.3 Mev, the nuclear force on it must be attractive and much stronger inside the nucleus than the repulsive electrostatic force. Such a nuclear force produces a potential well inside the nucleus, and we infer that the curve representing the potential energy of the alpha particle has a shape similar to that shown in Fig. 36 – 5.*

The dashed line in Fig. 36 – 5 represents the kinetic energy of an alpha particle at infinity. On the basis of the Newtonian mechanics this alpha particle cannot get closer than about twice the nuclear radius if it starts from the outside. By the same token, an alpha particle of this energy inside the potential well would never be able to get outside. We say there is a potential barrier holding the particle inside or outside as the case may be. Classically we expect no alpha decay. This will remain the expectation as the kinetic energy is increased until it reaches a value equal to or greater than the barrier height. This is not at all what happens, as we have seen from Fig. 36 – 4. Newtonian mechanics does not explain alpha decay at all. In fact, for all the cases in Fig. 36 – 4 it forbids it!

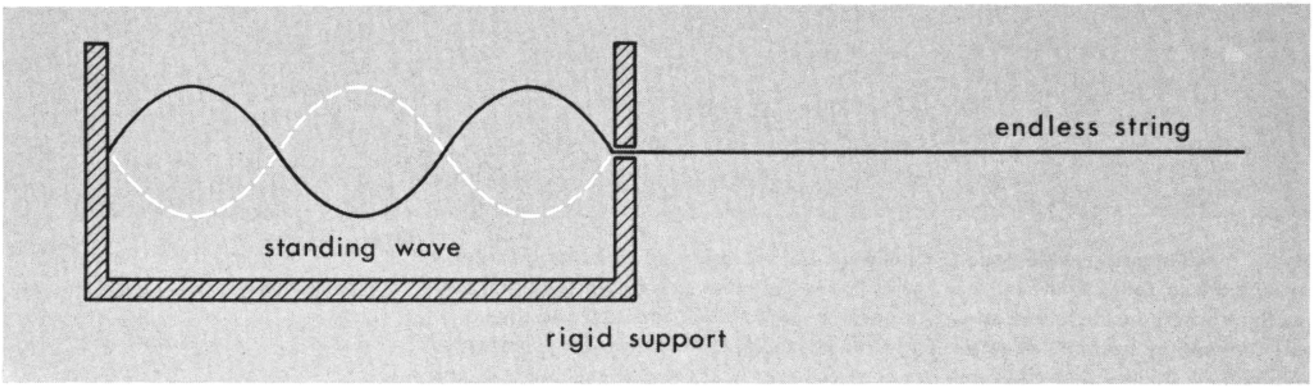
---

*When we make this inference we are starting on a *model* of alpha decay. Actually, the situation is certainly more complicated than we are assuming it to be. After all, the alpha particle is itself made up of four nucleons, and when it gets into the nuclear interior there is no guarantee that it will not get broken up by the strong nuclear forces there. The description would then have to be much more complicated. However, the results of our model are actually useful in making predictions about alpha decay, and are also illuminating in a general way.

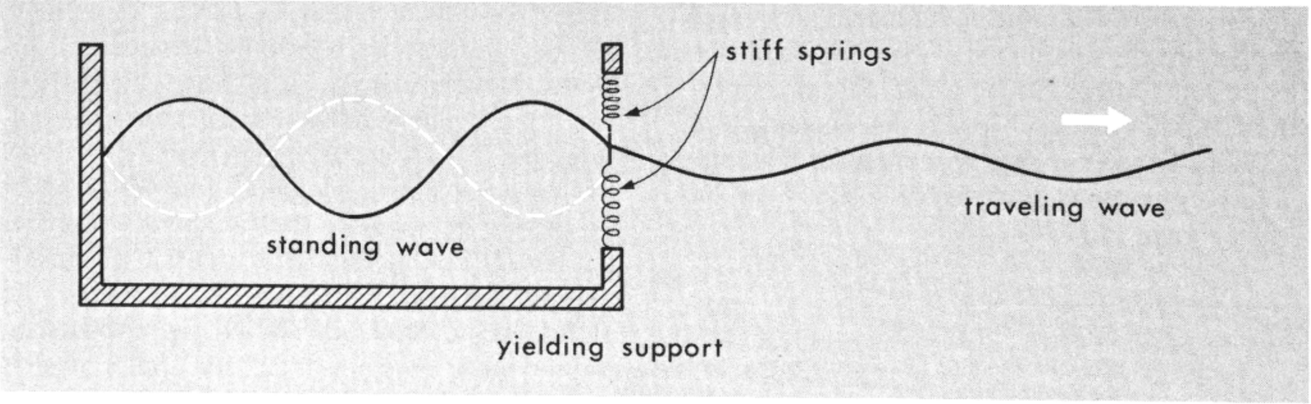
So we shall turn to our quantum wave model. It turns out that the wave analogy for the probability of finding particles is very effective in explaining the case of alpha decay, and this is the reason we have chosen to concentrate on alpha decay. But before we can really start our discussion, it will be necessary to go back to familiar wave systems in order to discuss the analogous effects there.

36–6 CLASSICAL WAVE MODELS. LIFETIMES AND BARRIER PENETRATION

Suppose we now consider a standing wave on a string that is attached to a rigid support, passes through a narrow hole in another rigid support, and extends endlessly beyond (Fig. 36 – 6). If we idealize away the effects of internal friction and of air resistance, the wave will remain between the supports forever, oscillating steadily in the required shape. The endless string beyond the hole will remain motionless, because the narrow hole prevents any up-and-down motion at point *A*. Now let us relax the conditions just a bit. In place of the hole, let the string be held at that point by a different support, quite strong and almost rigid, to be sure, but yielding a little under the varying pull of the attached vibrating string. As the standing wave vibrates on the short segment of string, the movable support will yield a little (Fig. 36 – 7). Its motion will in



36–6 A standing wave on a string stretched between two rigid supports. Even though the string comes through the hole in the support, no wave motion will be imparted to the endless string to the right of that point because the string cannot move up and down at the hole. It is rigidly held.



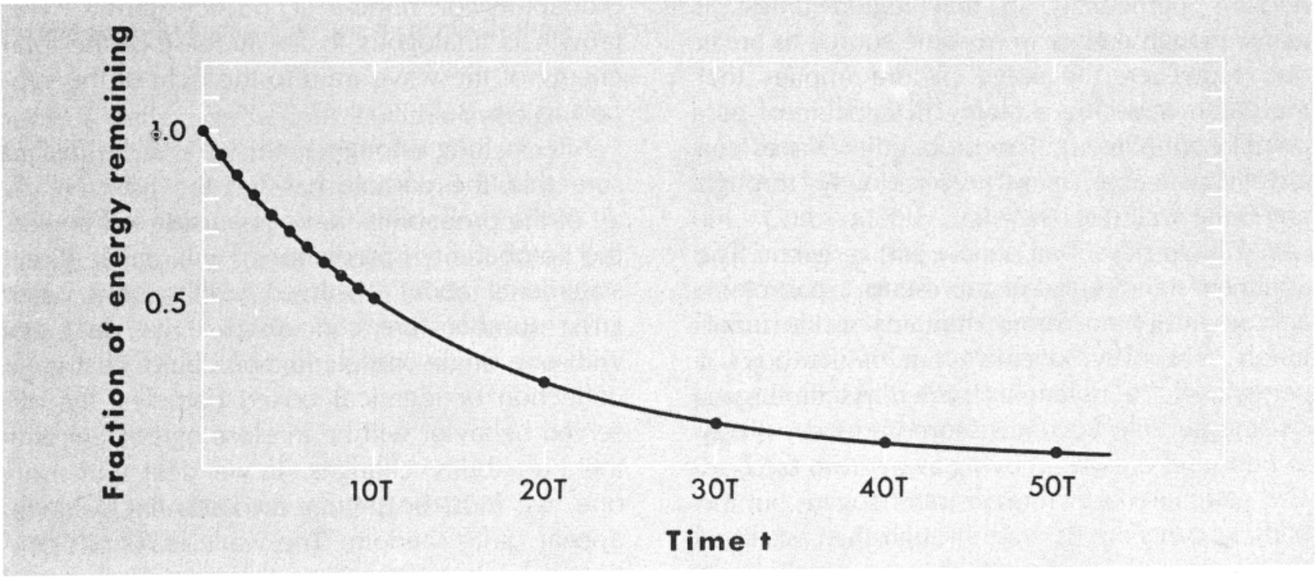
36–7 A standing wave for the case where one of the supports is not rigid. In this case the hole of Fig. 36 – 6 has been replaced by a pair of springs which allow the string to move up and down a little. Then a traveling wave is produced on the endless string by this motion. If the springs are very strong, the motion will be small, and so will the amplitude of the traveling wave.

turn impart a small motion to the long string beyond. Gradually some energy will be transferred from the standing wave between the narrow confines of its two supports into a vibration of the long string outside. A small traveling wave will be seen moving all the way out along the external string until finally all the motion within the original short segment has ceased. The wave has penetrated our support and the energy in it has leaked out!

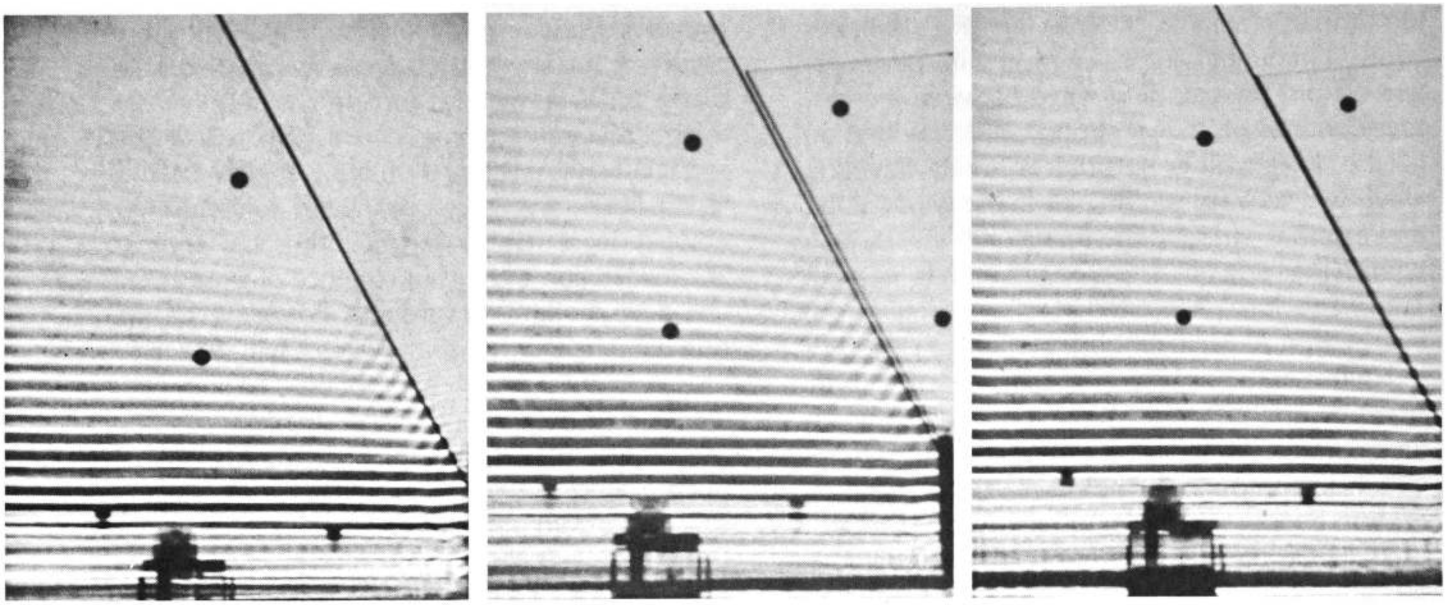
There are many ways to make a penetrable support. The support which can be used as an analogue for what happens in a nucleus is one which lets through a small but fixed fraction (f) of the energy of the standing wave per unit time. If we choose the period of vibration T as the unit of time, we can describe the "leaking" process as follows: Let the initial energy be E_0. Then after one vibration fE_0 will have leaked out, and $(1 - f)\ E_0$ is still stored in the standing wave. After the second vibration $(1 - f)^2E_0$ is still in the standing wave, and so on. If we plot the energy in the standing wave as a function of time we have a typical exponential decay curve (Fig. 36 – 8).

Penetration is seen in waves of every kind. The vibrations of a string will cause the most rigid possible support to yield a tiny bit. Light waves always penetrate a little way into the region beyond a surface from which they are totally reflected, and you can watch this take place in water waves, too. In the ripple tank, if you are careful, you can totally reflect the surface waves within a shallow region from a sharp boundary where the water suddenly becomes deep. Then, if you bring up another shallow region by pushing a piece of glass along the deep bottom toward the boundary of the original shallow portion, waves will begin to move off into the second shallow region as it approaches closer and closer to the first (Fig. 36 – 9). As you narrow the intervening gap of deep water, the waves that used to be totally reflected from the boundary will penetrate through it into the second shallow part of the tank. Finally, when the two shallow portions touch there will be no reflection.

Matter waves, which give the probability for finding particles, behave in a similar way. Leakage through a region where the potential energy is too great for motion from the point of view of Newtonian mechanics is always possible for the waves of probability. The flexible support in our string model and the deep water in our water wave model are the analogues of any potential which is not infinitely high or infinitely wide. In Newtonian physics, a potential barrier will confine a particle as long as it is high



36–8 The exponential decay of a standing wave. The time is plotted in units of the vibration period T. The dots represent the fraction of the initial energy after each vibration for the case when $(1 - f) = 0.933$, which gives $(1 - f)^{10} = 0.5$. The continuous curve is the limit when we consider the energy to be continuously leaked out.

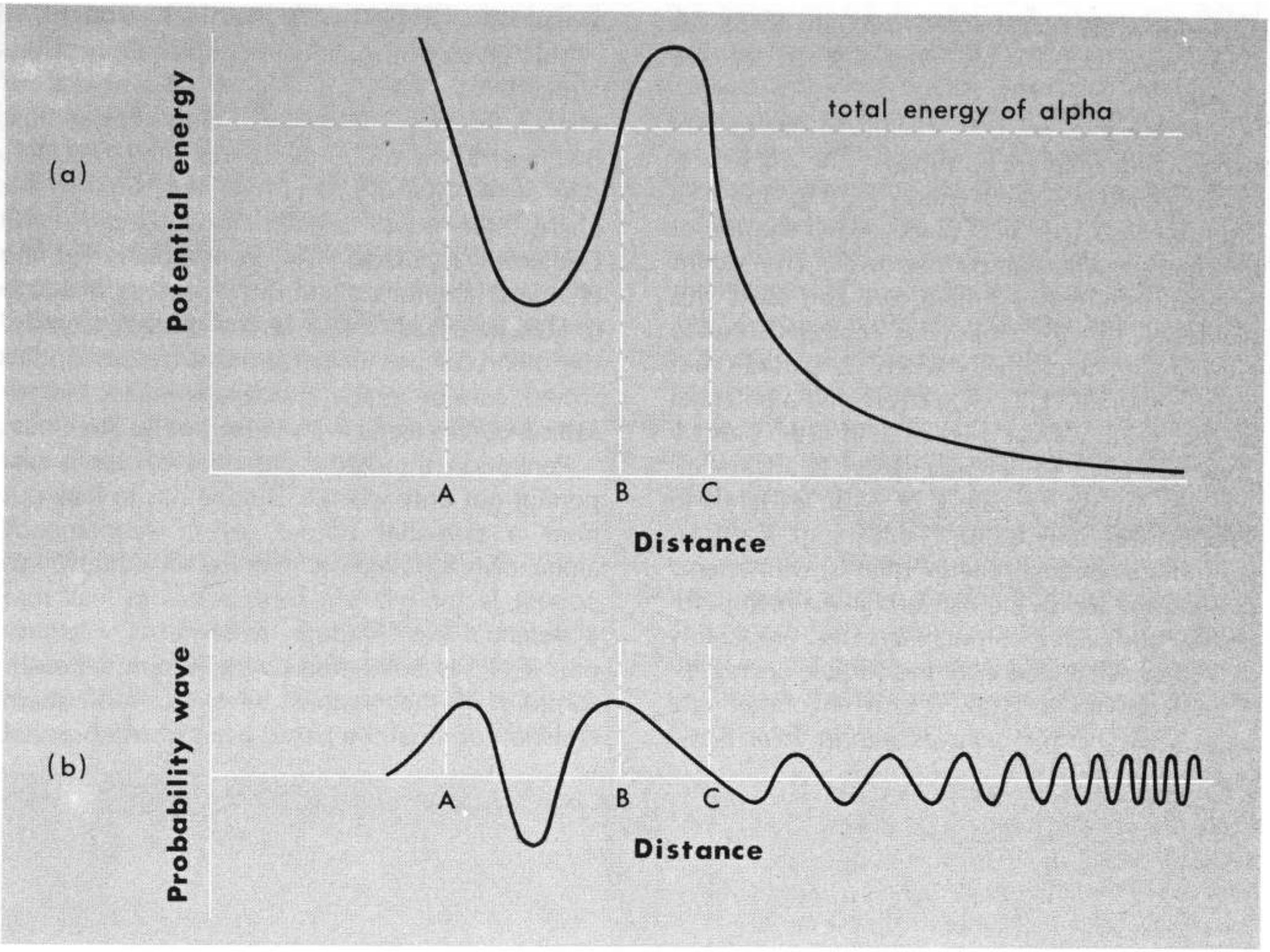


36–9 Barrier penetration of waves in a ripple tank. The black diagonal line in (a) is the edge of a glass plate that makes the water shallow to the left of it and deep to the right. The straight waves from the generator at the bottom of (a) are totally reflected from the boundary between the shallow and deep water. Notice that the waves penetrate a short distance into the deep water but do not continue in that medium. In (b) another glass plate has been moved close to the first one — about 1/4 wavelength away. Internal reflection still takes place, but some of the wave energy now leaks out and moves off across the shallow section on the right in spite of the narrow "barrier" of deep water between the plates. In (c) the plates are in contact and there is no internal reflection. The black dots in all three pictures are supports for the glass plates. (From the ESI film "Barrier Penetration.")

enough, no matter how narrow it is. The Newtonian result predicts that a particle, once confined by a potential to a certain region of space, remains perpetually in that region unless it gains enough energy from some source to break out. However, the wave picture implies that there can never be a real wall capable of perpetual confinement. The probability waves can always penetrate, if only very slowly, through any finite wall (Fig. 36 – 10).

We have described a new sort of event. The confined standing wave represents a particle in a box. But it no longer remains inside indefinitely. The wave "leaks" out of the box — it penetrates the potential barrier. As time goes on, the particle becomes more and more likely to be found outside, moving away from the box. The particle energy remains unchanged, but the particle can find its way through the "wall" of the box, to move off with a kinetic energy such that the total energy of the system remains conserved. We cannot predict exactly when we can observe the particle outside the box. The wave intensity inside steadily falls, and the probability of finding the particle there falls. But since the particle must be found somewhere, the probability of finding it outside grows. This growth is analogous to an increase of the total energy of the wave train to the right of the support in Fig. 36 – 7.

After a long enough time, we can be all but sure that the particle has left the box. Nearly all of the probability wave is outside. Of course, the probability predictions will give direct statements about observed results only when large numbers are concerned. If we deal, not with one single particle in a box but with a great collection of identical boxed particles, the observed behavior will be in close agreement with the probability changes. If we deal with only one, we must be prepared to see the behavior appear quite random. The wave leaks out continuously, but sometimes the particle is found to leave the box almost immediately; sometimes, instead, it will stay within for a long time. A small sample decays erratically; a single



36–10 The penetration of probability waves through a potential barrier which, according to Newtonian dynamics, reflects the particle. In the region between B and C, the probability wave leaks through the barrier. In the region to the left of A, the wave is also penetrating, but the potential continues to rise to the left of A so that the probability of finding a particle there rapidly drops to zero.

nucleus defies prediction: but the average behavior of many small samples, or the actual decay of a single large sample, will always approach the steady exponential decline plotted in Fig. 36 – 8. In our string model the half-life of the decay depended on the details of the leaky support. For a probability wave the lifetime will depend on the height and width of the potential barrier.

36–7 ALPHA DECAY AND BARRIER PENETRATION

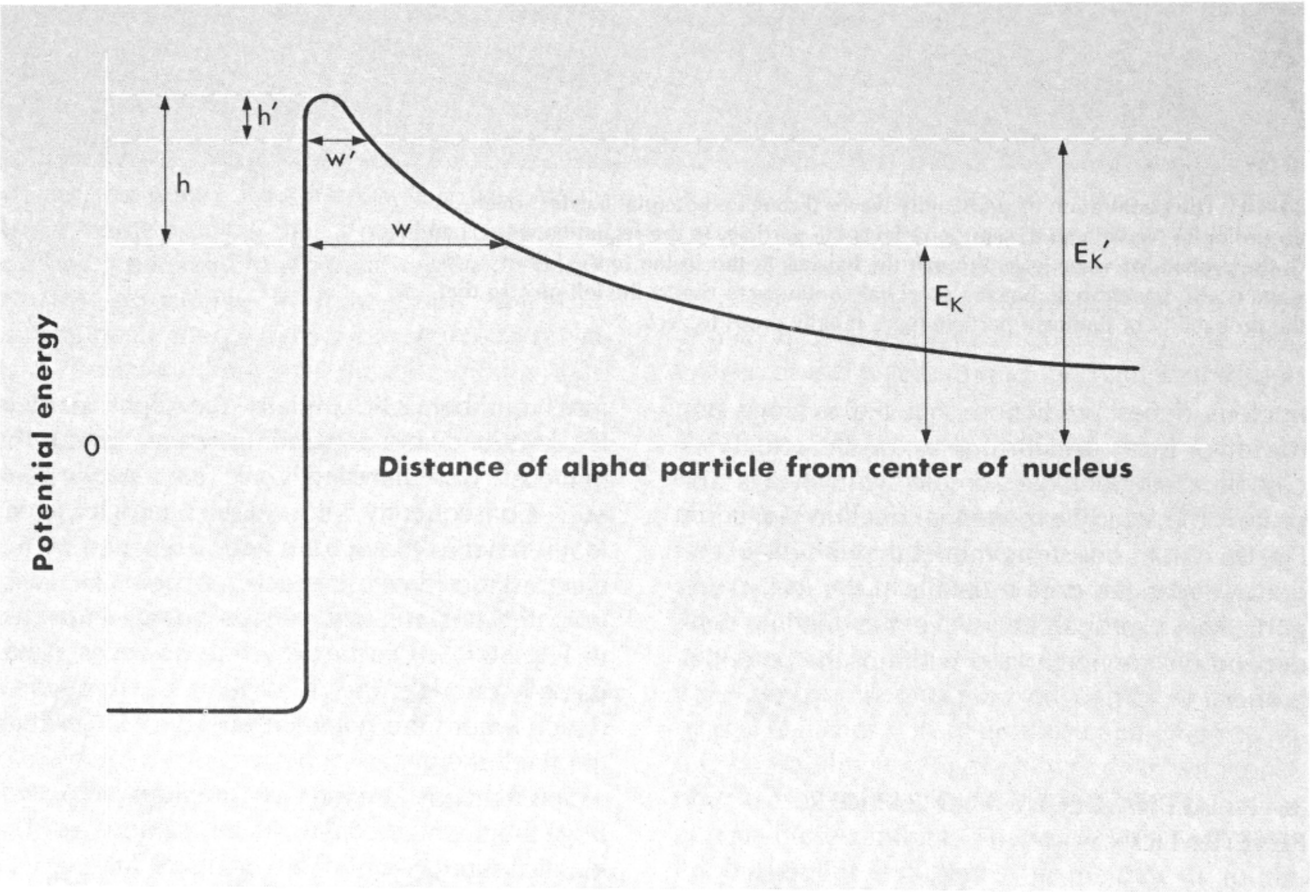
Let us now get back to our problem of alpha decay, say the decay of a uranium 238 nucleus into an alpha particle and a nucleus of atomic mass number 234. Initially the alpha particle is described by a standing wave inside the nucleus. This standing wave can slowly leak out. Consequently, the alpha particle eventually reaches the outside and is repelled by the electrostatic forces. It reaches a point far away from the nucleus with kinetic energy indicated in Fig. 36 – 5. On the way out, however, it has passed through the hill—that is, through a region where the potential energy is larger than the total energy.

Spontaneous fission of uranium into two large fragments behaves in the same way. The kinetic energy of the fragments is large when they are far apart; but as the fragments separate they pass through a region where the potential energy is greater than the total energy. Here again the waves leak slowly through. So fission takes place, but at a rate controlled by leakage through a barrier.

Now we can see how excitation energy can make a big difference. Suppose the nucleus is in an excited state with energy E'_K (Fig. 36 – 11). Then the leakage takes place nearer the top of the barrier where it is narrower. The waves penetrate more easily here. We can make this plausible by the following comparison. In Newtonian mechanics, when a particle has sufficient kinetic energy to overcome a potential barrier it will always pass over it. If the particle does not have sufficient energy, it always rebounds. The wave nature of particles tends to smooth out this abrupt transition. Particles with almost enough energy (that is, with energy close to the top of the barrier) have a high probability of leaking through it. This probability decreases as the energy of the particle decreases. A particle with very low energy encountering a high, wide barrier behaves almost in a Newtonian way—it will hardly ever leak through.

This gives us a qualitative explanation of the alpha-decay systems of Fig. 36 – 4. For a fixed charge, the alpha decays with more energy must be closer to the top of the barrier, and thus their rate of leakage is faster. Also, the smaller the charge, the lower is the barrier, because the Coulomb repulsion is proportional to the charge. Thus, for a fixed decay energy, the rate of leakage is faster the lower the charge. Actually, the theory of barrier penetration by probability waves can be made quantitative and the resulting curves agree with those in Fig. 36 – 4.

Pentration through a potential barrier is important not only when a particle has to leak out from a potential barrier, as in spontaneous alpha disintegration: it may be an equally important factor when a particle has to leak into a potential well through a barrier. For example, if N^{15} is bombarded with protons, the absorption of the protons with the subsequent emission of an alpha particle has been observed



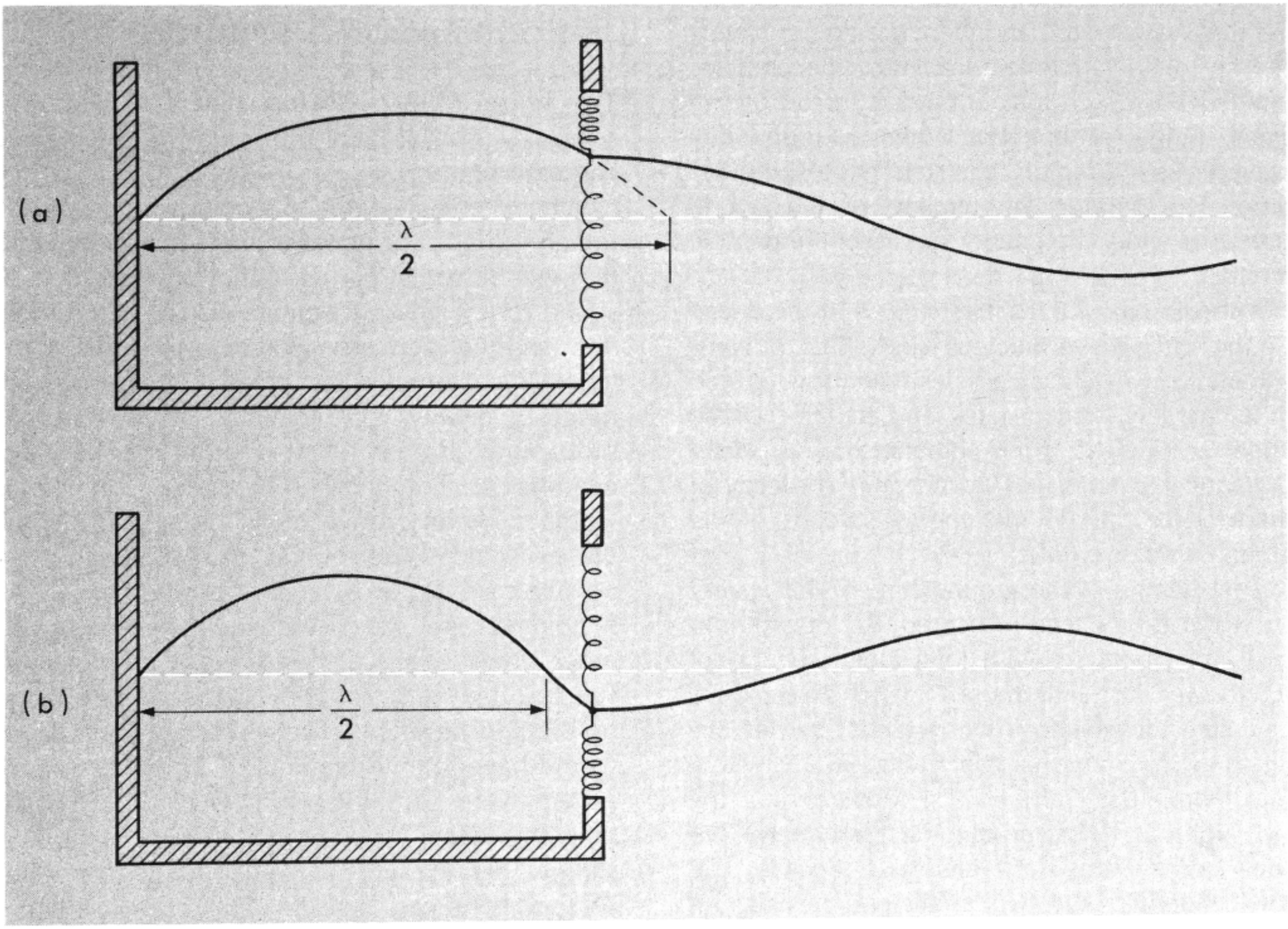
36–11 This figure is the same as Fig. 36–5, except that the excited-state energy E'_K has been put on the curve for comparison. Notice that w′, the width of the barrier, is less than w, the width for the unexcited nucleus. In addition, the energy h′ from the top of the barrier is less than that for the unexcited nucleus, where it is h. Both of these factors control the rate of barrier penetration.

by the use of protons with a kinetic energy of about 0.35 Mev. Now the height of the Coulomb potential barrier for the penetration of a proton into an N^{15} nucleus is about 1.5 Mev and, according to Newtonian mechanics, a proton of 0.35 Mev could not get closer than four times the nuclear radius. The observed reaction with the emission of an alpha particle is evidence that the proton wave manages to enter the nitrogen nucleus through the potential barrier. Similar barrier penetration is a very important factor in determining nuclear fusion reactions, which take place in the stars or may take place in controlled fusion.

36–8 LIFETIME AND ENERGY SPREAD

Between fixed, rigid supports, the standing waves of our string model can have only precisely fixed values for the wavelength: twice the length between supports, exactly the length, $\frac{2}{3}$ the length, $\frac{2}{4}$ the length, and so on: $\lambda = 2L/n$, with n any integer. In the same way, a quantum version with rigid walls gives only certain precise values for the wavelength of the standing probability waves, and therefore only certain precise energies for the particle they represent.

Now consider the more realistic model of a string with slightly yielding supports. Here the wavelength of the standing wave can vary a little, because the node need not be located exactly at the support. Figure 36 – 12 illustrates the situation. The string can contain a standing wave with a wavelength anywhere in the range shown in the figure. It can, therefore, also contain a standing wave which is a superposition of all possible standing waves in regions of allowed wavelengths. We have chosen a support which lets through a fixed fraction of the energy inside. Thus for larger amplitude between the supports, the support yields more, allowing the same fraction of the standing wave energy to escape. In this way, the range of allowed wavelengths



36–12 These two drawings show how a flexible support allows a range of wavelengths for a standing wave on a string between two supports. The wavelength in (a) is larger and the one in (b) is smaller than that obtained with a rigid support. Any wavelength between these two will be possible.

for the standing wave will be independent of the energy of the waves between the supports.

We see then, that a "leaky" support has two consequences. First, the standing wave decays, giving its energy to a traveling wave. Secondly, the standing wave can be made up of a superposition of many wavelengths near the precise values allowed by a rigid support. The same thing will therefore be true of the traveling wave. In this vibrating-string model, the less rigid the support, the greater the spread in wavelength the traveling wave will have. Also, the less rigid the support, the faster the wave will leak out. More concisely, the faster the wave leaks out, the greater is its spread in wavelength. A mechanical system which exhibits this behavior is described in detail in the Appendix to this chapter.

The same relations carry over to the quantum case. Whenever the states are only approximately stationary—that is, whenever they can decay by leakage of some particle—the particles which leak out cannot correspond to one sharply defined wavelength, and therefore do not have a sharply defined momentum. Also, the particles emitted from such a state have a spread over a small range of energy. When a probability wave leaks out rapidly, the state has a short lifetime. The shorter the lifetime of the bound state, the wider the energy spread of the particle emitted.

An example of this effect is seen in the decay of the radioactive nucleus $_4Be^8$. This nucleus splits into two alpha particles when it is formed in its first few excited states, and the effect of the width and height of the potential barrier is very dramatically illustrated in the rates of decay of these states and in the energy spreads of the alpha particles emitted.

The lifetime of the ground state of Be^8 is very short but it has been measured. Be^8 decays into two alpha particles with total kinetic energy of 0.096 Mev in a little more than 10^{-16} sec. If we calculate the height of the potential barrier, we again get a picture like that in Fig. 36 - 5, with a maximum height around 1.4 Mev, so that the two alpha particles can bind each other at 0.096 Mev. When this unstable state decays, the total energy of the two alpha particles has an uncertainty of about 4 ev. That is to say, individual measurements would give different answers, distributed around the value of 0.096 Mev with an energy spread of 4 ev. The figure 4 ev has been found by studying the formation of this bound state in the $\alpha - \alpha$ scattering experiments. In contrast, the first excited state of Be^8, about 2 Mev higher, has an energy spread of 2 Mev. It has a much shorter lifetime—too short to have been measured directly. In accordance with our expectation we find here that the shorter the lifetime, the less sharp is the emitted energy.

These two decays are examples of a very general relationship between lifetime and energy spread which is a consequence of the wave picture. This relation is

$$\tau \Delta E \approx h$$

where τ is the lifetime, ΔE is the energy spread of the emitted particles, and h is Planck's constant. Once this relation is known, it provides a way of measuring lifetimes so short that no known apparatus could "count the ticks." This indirect method carries our time scale down to the shortest known times. Here we measure not time directly but energy spread.

36–9 THE EMISSION OF PHOTONS

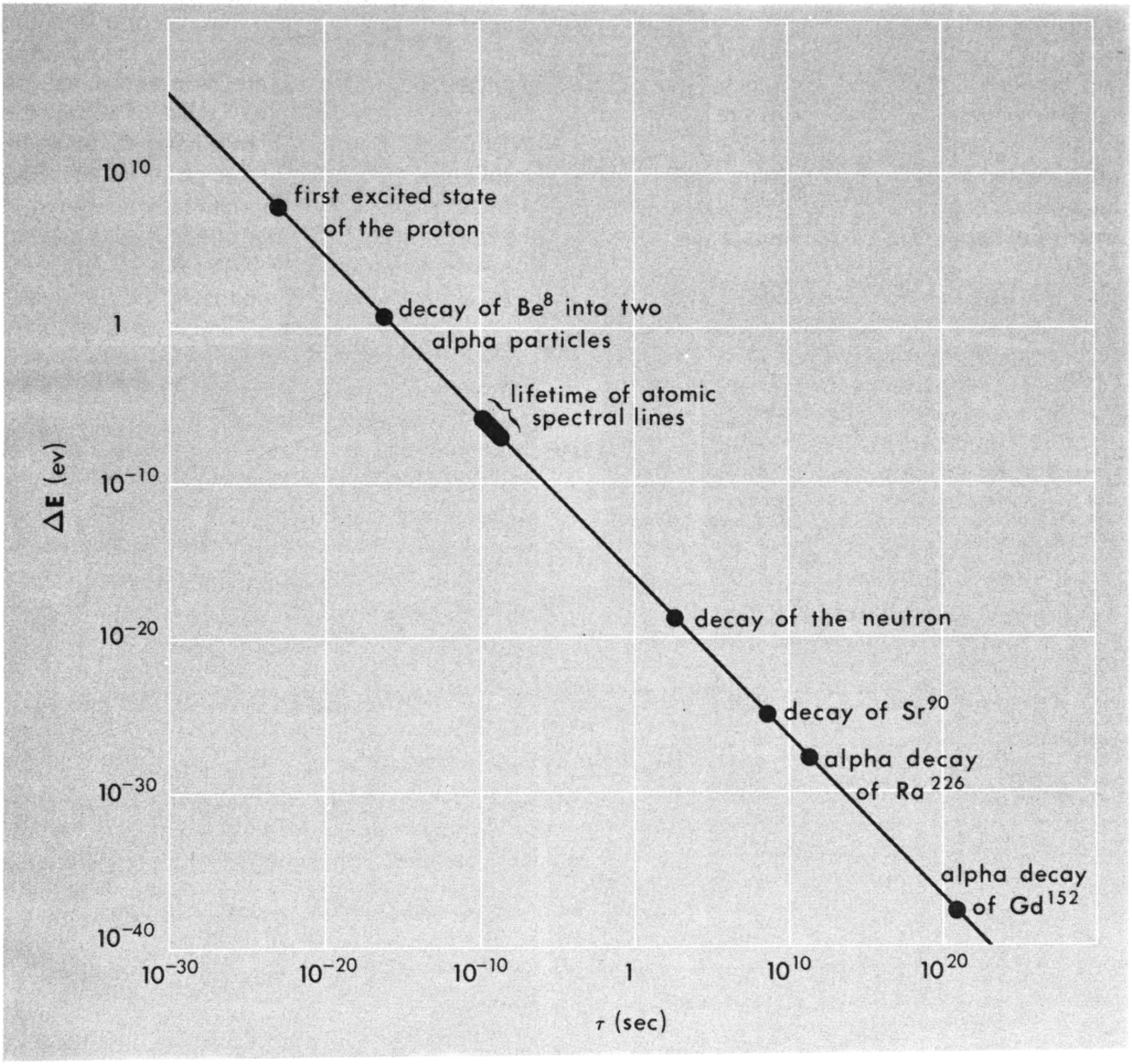
This brings us back to the subject of photons. Most systems—atomic, molecular, or nuclear—can emit photons.

Whenever states are only approximately stationary—that is, whenever they can decay by leakage of some particle—they cannot correspond to one sharply defined energy. The same is true when we have a state which emits photons. The photons also spread out into a little range of energy, or frequency. There is some probability that an excited atom can decay, emitting a photon with an energy anywhere within a narrow range of energy spread about the expected value for the ideal, perfectly stationary case. What is more, the relation $\tau \Delta E \approx h$ holds as well for photon emission as for all other types of decay. This means that every spectral line must have some spread of energy or it could never be observed at all, because it would have infinite lifetime. Only if the atomic system were bottled up forever, without possibility of leakage or emission, could it have one sharply defined energy value. Only one state of the atom can have a really sharp energy value: the ground state. For that state does not decay.

The line widths—the frequency spread of a single sprectrum line—can be observed optically with the best spectroscopic methods. Other effects—for instance, the Doppler shift in the observed frequencies caused by the random thermal motion of the different emitting atoms—usually mask the effect of the finite lifetime of the emitting state. But under the best conditions that atomic lifetime can be observed optically. This "natural width," as it is called, can be observed and usually amounts to about one part in 10^7 of the energy which itself is in the range of electron volts. On the other hand, a typical excited atom lasts on the average about 10^{-8} second before emitting an optical photon. We see here again our relationship $\tau \Delta E \approx h$.

The light wave oscillates in that time through about 10^7 cycles, so the "wave train" emitted by a single atom extends over a distance of about a meter. Somewhere within that train of waves, but not elsewhere, the photon can be found to induce its effect on the atoms of any sort of detector: eye, film, or photocell.

In the emission of photons by other systems, by molecules or by nuclei, for example, the lifetimes may be very different in order of magnitude. There are excited states of some nuclei which live years before emitting their photons—and, of course, the energy spread is correspondingly narrow! In Fig. 36–13 we have



36–13 Lifetimes of decaying states. In order to cover the enormous range of known lifetimes and line widths we have used logarithmic scales. Various decays discussed in the text have been put on the graph, ranging from the longest (narrowest) to the shortest (widest) presently known in physics. Most atomic spectral lines fall into the band shown in the figure, but there are also lines which fall well outside this band. We see that α decays span the whole figure from Gd to Be.

plotted ΔE vs τ on a log plot for all kinds of decays and crudely marked out the regions where the various types of decay usually fall. We can see how enormous the range is, how the lifetimes of some systems can be investigated only by measuring ΔE, and on the other extreme how the energy spread of systems can be investigated only by measuring τ.

APPENDIX

The Effect of Leakage on the Sharpness of Energy Levels: An Analogue Demonstration

In this Appendix we shall discuss a mechanical model which can be used to show the relationship between lifetime and frequency spread for standing waves. Torsional waves are generated by the to-and-fro twisting of a vertical steel wire in the center of Fig. A, the restoring force being the resistance of the wire to twist. Long transverse rods tipped with white discs are attached to the wire a few centimeters apart as loading masses.

The rods are made to oscillate by a periodic twisting of the upper end of the wire by a motor drive (not shown in the picture), and the resultant motion is photographed with a backdrop arranged as in Fig. B so that only the white discs are visible. The lower end of the steel wire is attached to a pair of paddles, perforated with small holes, which move in water as the wire twists. This termination has been carefully designed and adjusted to prevent wave reflections from the lower end of the wire. Thus the wire behaves as though it were infinitely long.

Our first experiment with this device is arranged to show how the leakage of a standing wave is related to its frequency spread. The photographs in Fig. C are time exposures, on which the discs produce white streaks; the length of the streak made by each disc shows the amplitude of vibration at that point on the wire. A meter stick mounted horizontally as shown in the upper left-hand corner of Fig. A was used to measure the amplitude.

As a standard for comparison, the "no leak"

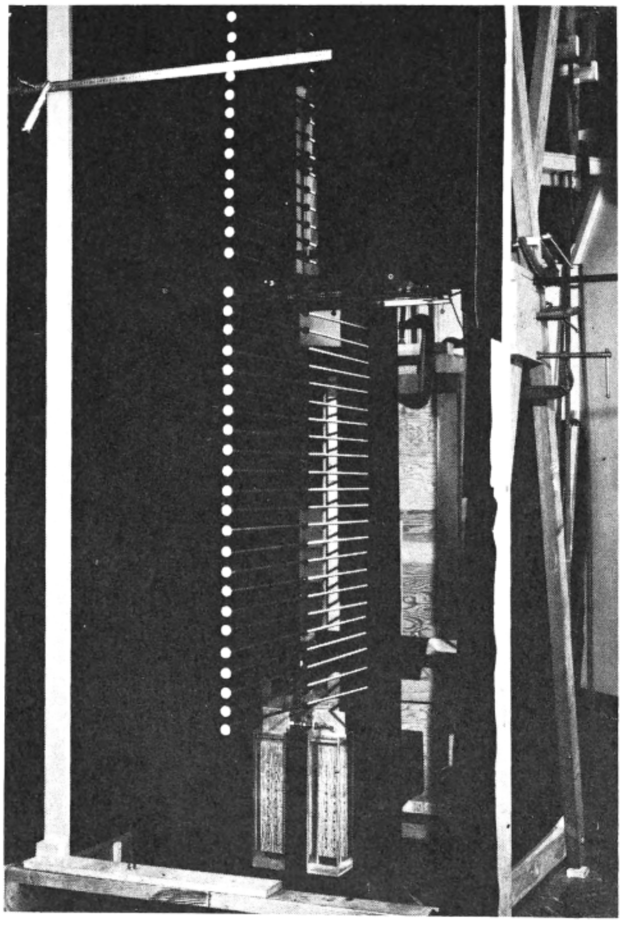


Figure A

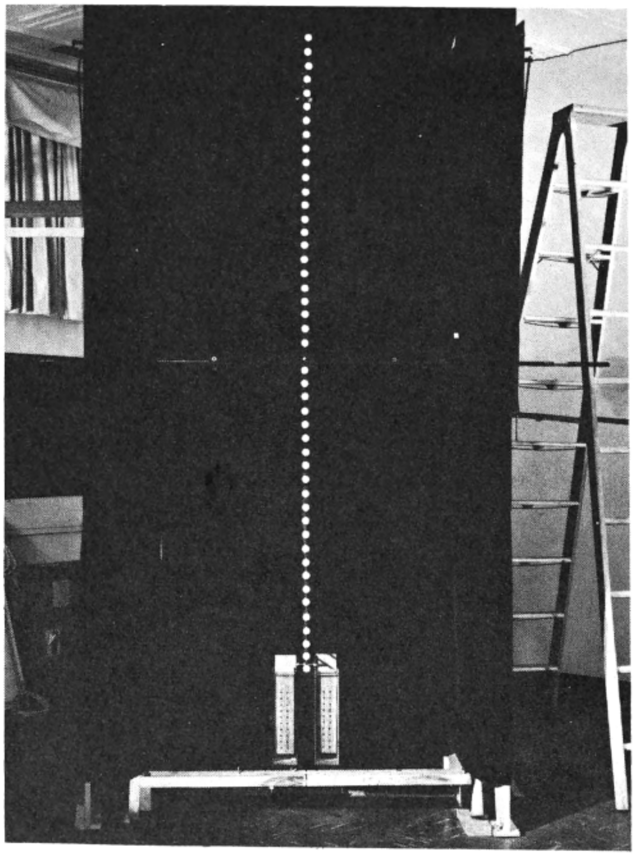
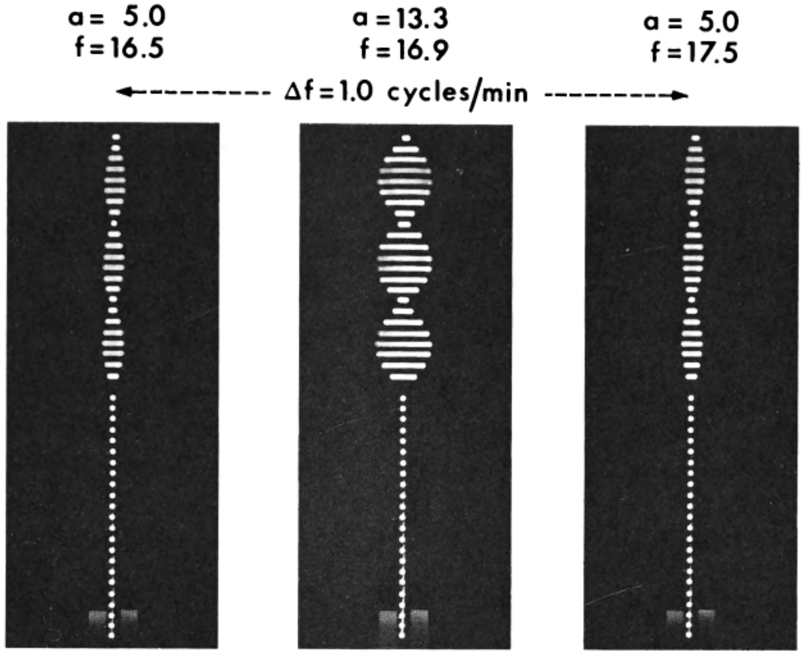


Figure B

case is shown in the three upper photographs of Fig. C. One of the rods along the wire (the twenty-fourth from the top) is fixed securely so that the wire cannot twist at all at that point; this is the analogue of the rigid support of Fig. 36 – 6. This locked rod has no white disc and shows up in the photographs of Fig. C as a gap in the rows of white dots. When the wire is twisted periodically, a standing wave is produced between the driver and the locked rod. As the frequency *f* of twist is varied continuously, the amplitude *a* of the standing wave varies. The

NO LEAK


a = 5.0
f = 16.5
a = 13.3
f = 16.9
a = 5.0
f = 17.5
Δf = 1.0 cycles/min

WITH LEAK

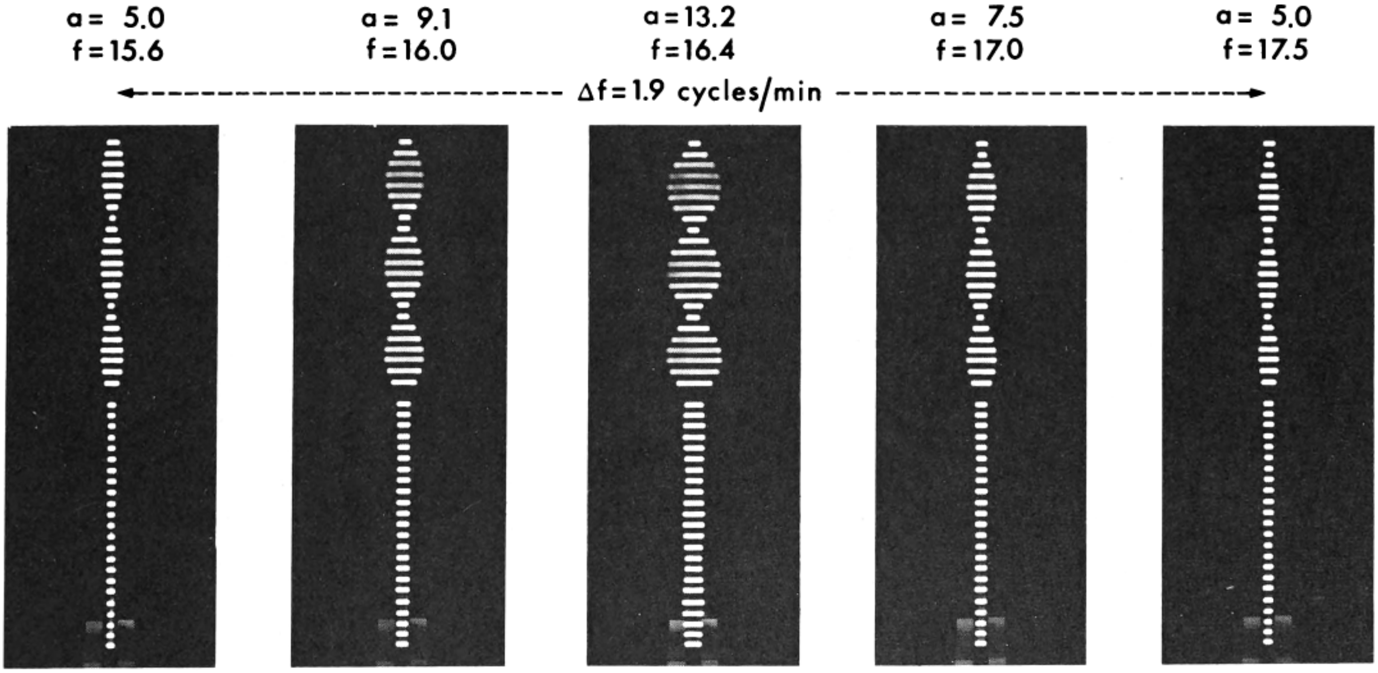
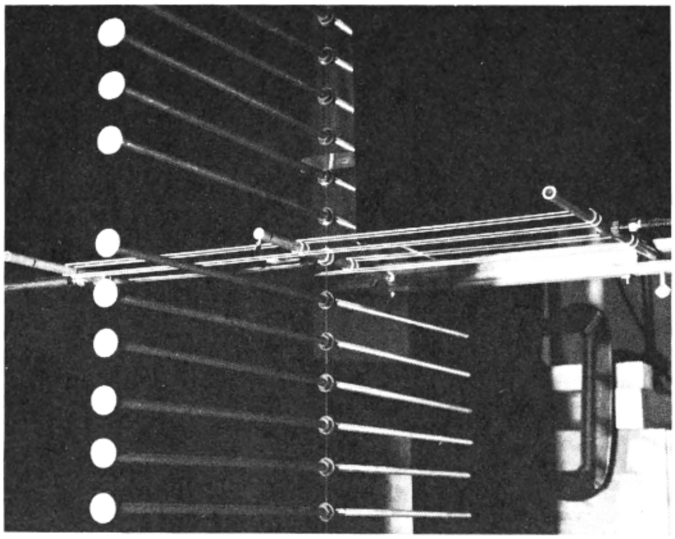

a = 5.0
f = 15.6
a = 9.1
f = 16.0
a = 13.2
f = 16.4
a = 7.5
f = 17.0
a = 5.0
f = 17.5
Δf = 1.9 cycles/min

Figure C

amplitude (shown in cm) is appreciable over the frequency range between 16.5 and 17.5 cycles per minute (cpm) and reaches its maximum near the midpoint of this range, where it corresponds to the standing wave with two nodes between the driver and the locked rod. The width of the frequency band which gives an appreciable amplitude is therefore about $\Delta f = 1.0$ cycle/min. If the system were ideal, there would be an appreciable amplitude at practically only one frequency: there would be no band. The observed frequency spread is due to frictional losses in the system.

The photographs in the lower part of Fig. C show the case in which we deliberately introduce an appreciable leak of the standing wave out of the upper part of the wire. This is done by partially freeing the locked rod. Instead of being rigidly held, it is restrained by elastic bands as shown in Fig. D. Once again the twisting frequency is varied, and it can be seen from the photographs that the range of frequency from one that gives an appreciable but small amplitude, through the maximum amplitude, to a similar small amplitude is wider than in the case where there was no leak. This frequency range is now $\Delta f = 1.9$ cycle/min. The traveling wave produced by the leak has the same spread of frequencies as the standing wave.

Now that we have seen how the frequency spread is related to leakage, we turn to the question of the lifetime of the standing wave. We investigate this by driving the system at the frequency which produces the maximum amplitude and then turning off the driver. When the twenty-fourth rod is locked—that is, when there is no leak—the standing wave does not remain forever, because of the frictional losses previously mentioned. Nevertheless, a comparison shows that the amplitude of the standing wave decreases faster in the leaking system than in the nonleaking one. For example, in one test with the apparatus described here, it took only about 0.65 as long for the standing wave to lose half its original amplitude with the leak as it did without the leak. Ideally, the standing wave should go on forever if there is no leak.



QUESTIONS

1.* An oxygen molecule is in the excited state shown in Fig. 36 – 1(a). If it dissociates, what is the total kinetic energy of the two separated atoms?

2.* Li^6 has an excited nuclear energy level at 3.6 Mev above the ground state. Where on Fig. 36 – 1(b) would you draw this level?

3. (a) What is the order of magnitude of the temperature at which the average kinetic energy of oxygen molecules is comparable with the dissociation energy?

   (b) Why is the dissociation observed at considerably lower temperatures?

4.* A positron and an electron in free space can annihilate each other with all their rest energy taken up by two photons. Why is it impossible for all the energy to go into one photon?

5. Calculate the angular momentum of the electron in the wave theory of the hydrogen energy levels. In the transition from one level to its neighbor, how much angular momentum must the photon carry away if angular momentum is to be conserved? (Hint: recall that $\lambda = h/mv$ and $n\lambda = 2\pi r$.)

6.* A deuteron does not decay into a pair of free protons. Which conservation laws forbid this?

7. According to the conservation laws in this chapter, is it possible for a triton (H^3 nucleus) to decay into a He^3 nucleus plus an electron?

8. Is the reaction ${}_{18}Ar^{40} \rightarrow {}_{20}Ca^{40} + 2e^-$ possible? (The binding energies per particle for A^{40} and Ca^{40} are 8.5960 Mev and 8.5526 Mev, respectively.)

9.* Are the following nuclei arranged in order of increasing total binding energy: He^4, O^{16}, Mo^{96}, Ni^{60}?

10.* Find the mass m of a nucleus with mass number A, atomic number Z, and binding energy per particle B, in terms of the masses of the proton and neutron, m_p and m_n, respectively.

11.* Use Fig. 36 – 3 to show that Li^6 in its ground state cannot disintegrate into a deuteron (H^2) and an alpha particle (He^4).

12. Use the data from Table 2, Chapter 32 to calculate the binding energy per particle of He^4.

13. How many joules of kinetic energy could be shared by Zr^{90} and Nd^{146} nuclei resulting from the fission of 1 gm of U^{236} nuclei in the ground state?

14. In 1938, Bethe proposed the following sequence of nuclear reactions as a possible source of the energy of the sun:

$${}_6C^{12} + {}_1H^1 \rightarrow {}_7N^{13} + \gamma$$
$${}_7N^{13} \rightarrow {}_6C^{13} + e^+ + \nu$$
$$e^- + e^+ \rightarrow \gamma + \gamma$$
$${}_6C^{13} + {}_1H^1 \rightarrow {}_7N^{14} + \gamma$$
$${}_7N^{14} + {}_1H^1 \rightarrow {}_8O^{15} + \gamma$$
$${}_8O^{15} \rightarrow {}_7N^{15} + e^+ + \nu$$
$$e^- + e^+ \rightarrow \gamma + \gamma$$
$${}_7N^{15} + {}_1H^1 \rightarrow {}_6C^{12} + {}_2He^4$$

The net result of the cycle is the conversion of four hydrogen atoms into one helium atom, the production of numerous gamma rays and neutrinos (ν), and recoiling heavy nuclei. What is the kinetic energy of all the particles if one mole of hydrogen atoms reacts?

15.* From Fig. 36 – 4 which nucleus, Rn^{222} or Rn^{218}, has a greater probability of emitting an alpha particle during a given time interval?

16.* How does the kinetic energy of an alpha particle inside the nucleus, Fig. 36 – 5., compare with E_K, its kinetic energy at a large distance from the nucleus?

17. According to Newtonian mechanics, what is the classical distance of closest approach to a nucleus of ${}_{96}Cm^{240}$ for a 6.3 Mev alpha particle?

18. Which of the nuclei Po^{212}, Bi^{210}, and Pb^{208} are stable and which unstable against alpha-particle emission? Data for the nuclei are in the following table where B is the total binding energy.

Nucleus	*B*
${}_{84}Po^{212}$	1657 Mev
${}_{83}Bi^{210}$	1646 Mev
${}_{82}Pb^{208}$	1637 Mev
${}_{81}Tl^{206}$	1622 Mev
${}_{80}Hg^{204}$	1609 Mev
${}_2He^4$	28 Mev

19. Let us consider a string stretched through two yielding supports of the type shown in Fig. 36 – 7. If we set up a standing wave between the supports, how will its lifetime compare with the case of Fig. 36 – 7? How will the spreads in the wavelength compare?

20. Suppose a long string runs through the leaky support of Fig. 36 – 7. What will happen if a short pulse is incident on the support from the right?

21. A beam of light is incident on a plane glass-to-air interface at an angle such that it is totally reflected. Suppose we bring another piece of glass up to the interface. What do you expect will happen as the glass gets closer and closer? Would the result depend on the wavelength of the light?

22. (a) Suppose that a vibrating string of initial energy E_0 has a penetrable support so that after each vibration 0.10 of its energy has leaked out. After how many vibrations will the energy left in the standing wave be less than 0.75 E_0? less than 0.50 E_0? less than 0.33 E_0?

(b) Suppose the standing wave which gives the probability of finding a particle in a box "leaks" out at such a rate that the probability decreases by 0.1 of its initial value in 1 minute. What is the probability of finding the particle still in the box after 2 minutes? After 5 minutes?

(c) Suppose 1000 particles were put into identical boxes, all of which had the same "leak" of probability waves as in (b). How many would you expect to find still in boxes after 5 minutes?

(d) Of the 1000 particles put into the boxes, how many would you expect to leave them during the first minute? During the second minute?

23.* Calculate the order of magnitude of the product $\tau\Delta E$ for the ground state of Be^8 and compare with the value of h.

24. Estimate the relative width $\Delta E/E$ of the energy peak of the gamma ray emitted in the transition from the first excited state of the nucleus K^{41}. ($E = 1.47$ Mev, $\tau = 9.5 \times 10^{-9}$ sec.)

25. Suppose we have a beam of excited sodium atoms moving at a velocity of about 10^7 cm/sec. They can emit light of a wavelength of about 6000 Å. Looking at different points along the beam, we note that the intensity of the light emitted from the beam falls off with the distance along the beam so that it is reduced by a factor of about 2 in 1 mm. Estimate the lifetime of the excited atom and the natural line width.

26. The first excited state of Fe^{57} has a lifetime of 10^{-6} sec and emits a 1.5×10^4 ev gamma ray in transition to the ground state. Because of the recoil of the nucleus when the photon is emitted, the photon energy differs from the energy difference between the two nuclear states.

(a) By how much does the photon energy differ?

(b) Compare this with the energy spread of the excited state.

(c) Can such a photon be absorbed by another Fe^{57} nucleus at rest in its ground state?

APPENDIX

INDEX

ANSWERS TO SHORT QUESTIONS

APPENDIX 1 Table of Trigonometric Functions

sin (read down)

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9		
0°	.0000	.0017	.0035	.0052	.0070	.0087	.0105	.0122	.0140	.0157	.0175	**89°**
1°	.0175	.0192	.0209	.0227	.0244	.0262	.0279	.0297	.0314	.0332	.0349	**88°**
2°	.0349	.0366	.0384	.0401	.0419	.0436	.0454	.0471	.0488	.0506	.0523	**87°**
3°	.0523	.0541	.0558	.0576	.0593	.0610	.0628	.0645	.0663	.0680	.0698	**86°**
4°	.0698	.0715	.0732	.0750	.0767	.0785	.0802	.0819	.0837	.0854	.0872	**85°**
5°	.0872	.0889	.0906	.0924	.0941	.0958	.0976	.0993	.1011	.1028	.1045	**84°**
6°	.1045	.1063	.1080	.1097	.1115	.1132	.1149	.1167	.1184	.1201	.1219	**83°**
7°	.1219	.1236	.1253	.1271	.1288	.1305	.1323	.1340	.1357	.1374	.1392	**82°**
8°	.1392	.1409	.1426	.1444	.1461	.1478	.1495	.1513	.1530	.1547	.1564	**81°**
9°	.1564	.1582	.1599	.1616	.1633	.1650	.1668	.1685	.1702	.1719	.1736	**80°**
10°	.1736	.1754	.1771	.1788	.1805	.1822	.1840	.1857	.1874	.1891	.1908	**79°**
11°	.1908	.1925	.1942	.1959	.1977	.1994	.2011	.2028	.2045	.2062	.2079	**78°**
12°	.2079	.2096	.2113	.2130	.2147	.2164	.2181	.2198	.2115	.2233	.2250	**77°**
13°	.2250	.2267	.2284	.2300	.2317	.2334	.2351	.2368	.2385	.2402	.2419	**76°**
14°	.2419	.2436	.2453	.2470	.2487	.2504	.2521	.2538	.2554	.2571	.2588	**75°**
15°	.2588	.2605	.2622	.2639	.2656	.2672	.2689	.2706	.2723	.2740	.2756	**74°**
16°	.2756	.2773	.2790	.2807	.2823	.2840	.2857	.2874	.2890	.2907	.2924	**73°**
17°	.2924	.2940	.2957	.2974	.2990	.3007	.3024	.3040	.3057	.3074	.3090	**72°**
18°	.3090	.3107	.3123	.3140	.3156	.3173	.3190	.3206	.3223	.3239	.3256	**71°**
19°	.3256	.3272	.3289	.3305	.3322	.3338	.3355	.3371	.3387	.3404	.3420	**70°**
20°	.3420	.3437	.3453	.3469	.3486	.3502	.3518	.3535	.3551	.3567	.3584	**69°**
21°	.3584	.3600	.3616	.3633	.3649	.3665	.3681	.3697	.3714	.3730	.3746	**68°**
22°	.3746	.3762	.3778	.3795	.3811	.3827	.3843	.3859	.3875	.3891	.3907	**67°**
23°	.3907	.3923	.3939	.3955	.3971	.3987	.4003	.4019	.4035	.4051	.4067	**66°**
24°	.4067	.4083	.4099	.4115	.4131	.4147	.4163	.4179	.4195	.4210	.4226	**65°**
25°	.4226	.4242	.4258	.4274	.4289	.4305	.4321	.4337	.4352	.4368	.4384	**64°**
26°	.4384	.4399	.4415	.4431	.4446	.4462	.4478	.4493	.4509	.4524	.4540	**63°**
27°	.4540	.4555	.4571	.4586	.4602	.4617	.4633	.4648	.4664	.4679	.4695	**62°**
28°	.4695	.4710	.4726	.4741	.4756	.4772	.4787	.4802	.4818	.4833	.4848	**61°**
29°	.4848	.4863	.4879	.4894	.4909	.4924	.4939	.4955	.4970	.4985	.5000	**60°**
30°	.5000	.5015	.5030	.5045	.5060	.5075	.5090	.5105	.5120	.5135	.5150	**59°**
31°	.5150	.5165	.5180	.5195	.5210	.5225	.5240	.5255	.5270	.5284	.5299	**58°**
32°	.5299	.5314	.5329	.5344	.5358	.5373	.5388	.5402	.5417	.5432	.5446	**57°**
33°	.5446	.5461	.5476	.5490	.5505	.5519	.5534	.5548	.5563	.5577	.5592	**56°**
34°	.5592	.5606	.5621	.5635	.5650	.5664	.5678	.5693	.5707	.5721	.5736	**55°**
35°	.5736	.5750	.5764	.5779	.5793	.5807	.5821	.5835	.5850	.5864	.5878	**54°**
36°	.5878	.5892	.5906	.5920	.5934	.5948	.5962	.5976	.5990	.6004	.6018	**53°**
37°	.6018	.6032	.6046	.6060	.6074	.6088	.6101	.6115	.6129	.6143	.6157	**52°**
38°	.6157	.6170	.6184	.6198	.6211	.6225	.6239	.6252	.6266	.6280	.6293	**51°**
39°	.6293	.6307	.6320	.6334	.6347	.6361	.6374	.6388	.6401	.6414	.6428	**50°**
40°	.6428	.6441	.6455	.6468	.6481	.6494	.6508	.6521	.6534	.6547	.6561	**49°**
41°	.6561	.6574	.6587	.6600	.6613	.6626	.6639	.6652	.6665	.6678	.6691	**48°**
42°	.6691	.6704	.6717	.6730	.6743	.6756	.6769	.6782	.6794	.6807	.6820	**47°**
43°	.6820	.6833	.6845	.6858	.6871	.6884	.6896	.6909	.6921	.6934	.6947	**46°**
44°	.6947	.6959	.6972	.6984	.6997	.7009	.7022	.7034	.7046	.7059	.7071	**45°**
		.9	.8	.7	.6	.5	.4	.3	.2	.1	.0	

cos (read up)

Table of Trigonometric Functions

sin (read down)

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9		
45°	.7071	.7083	.7096	.7108	.7120	.7133	.7145	.7157	.7169	.7181	.7193	44°
46°	.7193	.7206	.7218	.7230	.7242	.7254	.7266	.7278	.7290	.7302	.7314	43°
47°	.7314	.7325	.7337	.7349	.7361	.7373	.7385	.7396	.7408	.7420	.7431	42°
48°	.7431	.7443	.7455	.7466	.7478	.7490	.7501	.7513	.7524	.7536	.7547	41°
49°	.7547	.7559	.7570	.7581	.7593	.7604	.7615	.7627	.7638	.7649	.7660	40°
50°	.7660	.7672	.7683	.7694	.7705	.7716	.7727	.7738	.7749	.7760	.7771	39°
51°	.7771	.7782	.7793	.7804	.7815	.7826	.7837	.7848	.7859	.7869	.7880	38°
52°	.7880	.7891	.7902	.7912	.7923	.7934	.7944	.7955	.7965	.7976	.7986	37°
53°	.7986	.7997	.8007	.8018	.8028	.8039	.8049	.8059	.8070	.8080	.8090	36°
54°	.8090	.8100	.8111	.8121	.8131	.8141	.8151	.8161	.8171	.8181	.8192	35°
55°	.8192	.8202	.8211	.8221	.8231	.8241	.8251	.8261	.8271	.8281	.8290	34°
56°	.8290	.8300	.8310	.8320	.8329	.8339	.8348	.8358	.8368	.8377	.8387	33°
57°	.8387	.8396	.8406	.8415	.8425	.8434	.8443	.8453	.8462	.8471	.8480	32°
58°	.8480	.8490	.8499	.8508	.8517	.8526	.8536	.8545	.8554	.8563	.8572	31°
59°	.8572	.8581	.8590	.8599	.8607	.8616	.8625	.8634	.8643	.8652	.8660	30°
60°	.8660	.8669	.8678	.8686	.8695	.8704	.8712	.8721	.8729	.8738	.8746	29°
61°	.8746	.8755	.8763	.8771	.8780	.8788	.8796	.8805	.8813	.8821	.8829	28°
62°	.8829	.8838	.8846	.8854	.8862	.8870	.8878	.8886	.8894	.8902	.8910	27°
63°	.8910	.8918	.8926	.8934	.8942	.8949	.8957	.8965	.8973	.8980	.8988	26°
64°	.8988	.8996	.9003	.9011	.9018	.9026	.9033	.9041	.9048	.9056	.9063	25°
65°	.9063	.9070	.9078	.9085	.9092	.9100	.9107	.9114	.9121	.9128	.9135	24°
66°	.9135	.9143	.9150	.9157	.9164	.9171	.9178	.9184	.9191	.9198	.9205	23°
67°	.9205	.9212	.9219	.9225	.9232	.9239	.9245	.9252	.9259	.9265	.9272	22°
68°	.9272	.9278	.9285	.9291	.9298	.9304	.9311	.9317	.9323	.9330	.9336	21°
69°	.9336	.9342	.9348	.9354	.9361	.9367	.9373	.9379	.9385	.9391	.9397	20°
70°	.9397	.9403	.9409	.9415	.9421	.9426	.9432	.9438	.9444	.9449	.9455	19°
71°	.9455	.9461	.9466	.9472	.9478	.9483	.9489	.9494	.9500	.9505	.9511	18°
72°	.9511	.9516	.9521	.9527	.9532	.9537	.9542	.9548	.9553	.9558	.9563	17°
73°	.9563	.9568	.9573	.9578	.9583	.9588	.9593	.9598	.9603	.9608	.9613	16°
74°	.9613	.9617	.9622	.9627	.9632	.9636	.9641	.9646	.9650	.9655	.9659	15°
75°	.9659	.9664	.9668	.9673	.9677	.9681	.9686	.9690	.9694	.9699	.9703	14°
76°	.9703	.9707	.9711	.9715	.9720	.9724	.9728	.9732	.9736	.9740	.9744	13°
77°	.9744	.9748	.9751	.9755	.9759	.9763	.9767	.9770	.9774	.9778	.9781	12°
78°	.9781	.9785	.9789	.9792	.9796	.9799	.9803	.9806	.9810	.9813	.9816	11°
79°	.9816	.9820	.9823	.9826	.9829	.9833	.9836	.9839	.9842	.9845	.9848	10°
80°	.9848	.9851	.9854	.9857	.9860	.9863	.9866	.9869	.9871	.9874	.9877	9°
81°	.9877	.9880	.9882	.9885	.9888	.9890	.9893	.9895	.9898	.9900	.9903	8°
82°	.9903	.9905	.9907	.9910	.9912	.9914	.9917	.9919	.9921	.9923	.9925	7°
83°	.9925	.9928	.9930	.9932	.9934	.9936	.9938	.9940	.9942	.9943	.9945	6°
84°	.9945	.9947	.9949	.9951	.9952	.9954	.9956	.9957	.9959	.9960	.9962	5°
85°	.9962	.9963	.9965	.9966	.9968	.9969	.9971	.9972	.9973	.9974	.9976	4°
86°	.9976	.9977	.9978	.9979	.9980	.9981	.9982	.9983	.9984	.9985	.9986	3°
87°	.9986	.9987	.9988	.9989	.9990	.9990	.9991	.9992	.9993	.9993	.9994	2°
88°	.9994	.9995	.9995	.9996	.9996	.9997	.9997	.9997	.9998	.9998	.9998	1°
89°	.9998	.9999	.9999	.9999	.9999	1.000	1.000	1.000	1.000	1.000	1.000	0°
		.9	.8	.7	.6	.5	.4	.3	.2	.1	.0	

cos (read up)

Table of Trigonometric Functions

tan (read down)

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9		
0°	.0000	.0017	.0035	.0052	.0070	.0087	.0105	.0122	.0140	.0157	.0175	**89°**
1°	.0175	.0192	.0209	.0227	.0244	.0262	.0279	.0297	.0314	.0332	.0349	**88°**
2°	.0349	.0367	.0384	.0402	.0419	.0437	.0454	.0472	.0489	.0507	.0524	**87°**
3°	.0524	.0542	.0559	.0577	.0594	.0612	.0629	.0647	.0664	.0682	.0699	**86°**
4°	.0699	.0717	.0734	.0752	.0769	.0787	.0805	.0822	.0840	.0857	.0875	**85°**
5°	.0875	.0892	.0910	.0928	.0945	.0963	.0981	.0998	.1016	.1033	.1051	**84°**
6°	.1051	.1069	.1086	.1104	.1122	.1139	.1157	.1175	.1192	.1210	.1228	**83°**
7°	.1228	.1246	.1263	.1281	.1299	.1317	.1334	.1352	.1370	.1388	.1405	**82°**
8°	.1405	.1423	.1441	.1459	.1477	.1495	.1512	.1530	.1548	.1566	.1584	**81°**
9°	.1584	.1602	.1620	.1638	.1655	.1673	.1691	.1709	.1727	.1745	.1763	**80°**
10°	.1763	.1781	.1799	.1817	.1835	.1853	.1871	.1890	.1908	.1926	.1944	**79°**
11°	.1944	.1962	.1980	.1998	.2016	.2035	.2053	.2071	.2089	.2107	.2126	**78°**
12°	.2126	.2144	.2162	.2180	.2199	.2217	.2235	.2254	.2272	.2290	.2309	**77°**
13°	.2309	.2327	.2345	.2364	.2382	.2401	.2419	.2438	.2456	.2475	.2493	**76°**
14°	.2493	.2512	.2530	.2549	.2568	.2586	.2605	.2623	.2642	.2661	.2679	**75°**
15°	.2679	.2698	.2717	.2736	.2754	.2773	.2792	.2811	.2830	.2849	.2867	**74°**
16°	.2867	.2886	.2905	.2924	.2943	.2962	.2981	.3000	.3019	.3038	.3057	**73°**
17°	.3057	.3076	.3096	.3115	.3134	.3153	.3172	.3191	.3211	.3230	.3249	**72°**
18°	.3249	.3269	.3288	.3307	.3327	.3346	.3365	.3385	.3404	.3424	.3443	**71°**
19°	.3443	.3463	.3482	.3502	.3522	.3541	.3561	.3581	.3600	.3620	.3640	**70°**
20°	.3640	.3659	.3679	.3699	.3719	.3739	.3759	.3779	.3799	.3819	.3839	**69°**
21°	.3839	.3859	.3879	.3899	.3919	.3939	.3959	.3979	.4000	.4020	.4040	**68°**
22°	.4040	.4061	.4081	.4101	.4122	.4142	.4163	.4183	.4204	.4224	.4245	**67°**
23°	.4245	.4265	.4286	.4307	.4327	.4348	.4369	.4390	.4411	.4431	.4452	**66°**
24°	.4452	.4473	.4494	.4515	.4536	.4557	.4578	.4599	.4621	.4642	.4663	**65°**
25°	.4663	.4684	.4706	.4727	.4748	.4770	.4791	.4813	.4834	.4856	.4877	**64°**
26°	.4877	.4899	.4921	.4942	.4964	.4986	.5008	.5029	.5051	.5073	.5095	**63°**
27°	.5095	.5117	.5139	.5161	.5184	.5206	.5228	.5250	.5272	.5295	.5317	**62°**
28°	.5317	.5340	.5362	.5384	.5407	.5430	.5452	.5475	.5498	.5520	.5543	**61°**
29°	.5543	.5566	.5589	.5612	.5635	.5658	.5681	.5704	.5727	.5750	.5774	**60°**
30°	.5774	.5797	.5820	.5844	.5867	.5890	.5914	.5938	.5961	.5985	.6009	**59°**
31°	.6009	.6032	.6056	.6080	.6104	.6128	.6152	.6176	.6200	.6224	.6249	**58°**
32°	.6249	.6273	.6297	.6322	.6346	.6371	.6395	.6420	.6445	.6469	.6494	**57°**
33°	.6494	.6519	.6544	.6569	.6594	.6619	.6644	.6669	.6694	.6720	.6745	**56°**
34°	.6745	.6771	.6796	.6822	.6847	.6873	.6899	.6924	.6950	.6976	.7002	**55°**
35°	.7002	.7028	.7054	.7080	.7107	.7133	.7159	.7186	.7212	.7239	.7265	**54°**
36°	.7265	.7292	.7319	.7346	.7373	.7400	.7427	.7454	.7481	.7508	.7536	**53°**
37°	.7536	.7563	.7590	.7618	.7646	.7673	.7701	.7729	.7757	.7785	.7813	**52°**
38°	.7813	.7841	.7869	.7898	.7926	.7954	.7983	.8012	.8040	.8069	.8098	**51°**
39°	.8098	.8127	.8156	.8185	.8214	.8243	.8273	.8302	.8332	.8361	.8391	**50°**
40°	.8391	.8421	.8451	.8481	.8511	.8541	.8571	.8601	.8632	.8662	.8693	**49°**
41°	.8693	.8724	.8754	.8785	.8816	.8847	.8878	.8910	.8941	.8972	.9004	**48°**
42°	.9004	.9036	.9067	.9099	.9131	.9163	.9195	.9228	.9260	.9293	.9325	**47°**
43°	.9325	.9358	.9391	.9424	.9457	.9490	.9523	.9556	.9590	.9623	.9657	**46°**
44°	.9657	.9691	.9725	.9759	.9793	.9827	.9861	.9896	.9930	.9965	1.000	**45°**
		.9	.8	.7	.6	.5	.4	.3	.2	.1	.0	

cot (read up)

Table of Trigonometric Functions

tan (read down)

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9		
45°	1.000	1.003	1.007	1.011	1.014	1.018	1.021	1.025	1.028	1.032	1.036	**44°**
46°	1.036	1.039	1.043	1.046	1.050	1.054	1.057	1.061	1.065	1.069	1.072	**43°**
47°	1.072	1.076	1.080	1.084	1.087	1.091	1.095	1.099	1.103	1.107	1.111	**42°**
48°	1.111	1.115	1.118	1.122	1.126	1.130	1.134	1.138	1.142	1.146	1.150	**41°**
49°	1.150	1.154	1.159	1.163	1.167	1.171	1.175	1.179	1.183	1.188	1.192	**40°**
50°	1.192	1.196	1.200	1.205	1.209	1.213	1.217	1.222	1.226	1.230	1.235	**39°**
51°	1.235	1.239	1.244	1.248	1.253	1.257	1.262	1.266	1.271	1.275	1.280	**38°**
52°	1.280	1.285	1.289	1.294	1.299	1.303	1.308	1.313	1.317	1.322	1.327	**37°**
53°	1.327	1.332	1.337	1.342	1.347	1.351	1.356	1.361	1.366	1.371	1.376	**36°**
54°	1.376	1.381	1.387	1.392	1.397	1.402	1.407	1.412	1.418	1.423	1.428	**35°**
55°	1.428	1.433	1.439	1.444	1.450	1.455	1.460	1.466	1.471	1.477	1.483	**34°**
56°	1.483	1.488	1.494	1.499	1.505	1.511	1.517	1.522	1.528	1.534	1.540	**33°**
57°	1.540	1.546	1.552	1.558	1.564	1.570	1.576	1.582	1.588	1.594	1.600	**32°**
58°	1.600	1.607	1.613	1.619	1.625	1.632	1.638	1.645	1.651	1.658	1.664	**31°**
59°	1.664	1.671	1.678	1.684	1.691	1.698	1.704	1.711	1.718	1.725	1.732	**30°**
60°	1.732	1.739	1.746	1.753	1.760	1.767	1.775	1.782	1.789	1.797	1.804	**29°**
61°	1.804	1.811	1.819	1.827	1.834	1.842	1.849	1.857	1.865	1.873	1.881	**28°**
62°	1.881	1.889	1.897	1.905	1.913	1.921	1.929	1.937	1.946	1.954	1.963	**27°**
63°	1.963	1.971	1.980	1.988	1.997	2.006	2.014	2.023	2.032	2.041	2.050	**26°**
64°	2.050	2.059	2.069	2.078	2.087	2.097	2.106	2.116	2.125	2.135	2.145	**25°**
65°	2.145	2.154	2.164	2.174	2.184	2.194	2.204	2.215	2.225	2.236	2.246	**24°**
66°	2.246	2.257	2.267	2.278	2.289	2.300	2.311	2.322	2.333	2.344	2.356	**23°**
67°	2.356	2.367	2.379	2.391	2.402	2.414	2.426	2.438	2.450	2.463	2.475	**22°**
68°	2.475	2.488	2.500	2.513	2.526	2.539	2.552	2.565	2.578	2.592	2.605	**21°**
69°	2.605	2.619	2.633	2.646	2.660	2.675	2.689	2.703	2.718	2.733	2.747	**20°**
70°	2.747	2.762	2.778	2.793	2.808	2.824	2.840	2.856	2.872	2.888	2.904	**19°**
71°	2.904	2.921	2.937	2.954	2.971	2.989	3.006	3.024	3.042	3.060	3.078	**18°**
72°	3.078	3.096	3.115	3.133	3.152	3.172	3.191	3.211	3.230	3.251	3.271	**17°**
73°	3.271	3.291	3.312	3.333	3.354	3.376	3.398	3.420	3.442	3.465	3.487	**16°**
74°	3.487	3.511	3.534	3.558	3.582	3.606	3.630	3.655	3.681	3.706	3.732	**15°**
75°	3.732	3.758	3.785	3.812	3.839	3.867	3.895	3.923	3.952	3.981	4.011	**14°**
76°	4.011	4.041	4.071	4.102	4.134	4.165	4.198	4.230	4.264	4.297	4.331	**13°**
77°	4.331	4.366	4.402	4.437	4.474	4.511	4.548	4.586	4.625	4.665	4.705	**12°**
78°	4.705	4.745	4.787	4.829	4.872	4.915	4.959	5.005	5.050	5.097	5.145	**11°**
79°	5.145	5.193	5.242	5.292	5.343	5.396	5.449	5.503	5.558	5.614	5.671	**10°**
80°	5.671	5.730	5.789	5.850	5.912	5.976	6.041	6.107	6.174	6.243	6.314	**9°**
81°	6.314	6.386	6.460	6.535	6.612	6.691	6.772	6.855	6.940	7.026	7.115	**8°**
82°	7.115	7.207	7.300	7.396	7.495	7.596	7.700	7.806	7.916	8.028	8.144	**7°**
83°	8.144	8.264	8.386	8.513	8.643	8.777	8.915	9.058	9.205	9.357	9.514	**6°**
84°	9.514	9.677	9.845	10.02	10.20	10.39	10.58	10.78	10.99	11.20	11.43	**5°**
85°	11.43	11.66	11.91	12.16	12.43	12.71	13.00	13.30	13.62	13.95	14.30	**4°**
86°	14.30	14.67	15.06	15.46	15.89	16.35	16.83	17.34	17.89	18.46	19.08	**3°**
87°	19.08	19.74	20.45	21.20	22.02	22.90	23.86	24.90	26.03	27.27	28.64	**2°**
88°	28.64	30.14	31.82	33.69	35.80	38.19	40.92	44.07	47.74	52.08	57.29	**1°**
89°	57.29	63.66	71.62	81.85	95.49	114.6	143.2	191.0	286.5	573.0	∞	**0°**
		.9	.8	.7	.6	.5	.4	.3	.2	.1	.0	

cot (read up)

APPENDIX 2 Physical Constants and Conversion Factors

PHYSICAL CONSTANTS

Avogadro's number: $N_0 = 6.02 \times 10^{23}$

Speed of light: $c = 2.99793 \times 10^{8}$ m/sec

Gravitational constant: $G = 6.670 \times 10^{-11} \dfrac{\text{m}^3}{\text{kg-sec}^2}$

Boltzmann's constant: $k = 1.3805 \times 10^{-23} \dfrac{\text{joule}}{{}^\circ\text{K}}$

Constant in Coulomb's law: $k = 2.306 \times 10^{-28} \dfrac{\text{newton-m}^2}{(\text{elem. ch.})^2} = 8.988 \times 10^{9} \dfrac{\text{newton-m}^2}{\text{coulomb}^2}$

Mass of electron: $m_e = 9.109 \times 10^{-31}$ kg

Mass of proton: $m_p = 1.672 \times 10^{-27}$ kg

Constant in Ampère's circuital law: $K = 2 \times 10^{-7} \dfrac{\text{newton}}{\text{amp}^2}$ (exact, by definition)

Planck's constant: $h = 6.626 \times 10^{-34}$ joule-sec $= 4.136 \times 10^{-15}$ ev-sec

CONVERSION FACTORS

1 atomic mass unit $= 1.66 \times 10^{-27}$ kg

1 electron volt $= 1.602 \times 10^{-19}$ joule

1 coulomb $= 6.242 \times 10^{18}$ elem. ch.

APPENDIX 3 Table of the Elements

ELEMENT	SYMBOL	ATOMIC NUMBER	ATOMIC MASS (in amu)
Actinium	Ac	89	(227.0278)
Aluminum	Al	13	26.9815
Americium	Am	95	(243.0614)
Antimony	Sb	51	121.75
Argon	Ar	18	39.948
Arsenic	As	33	74.9216
Astatine	At	85	(210)
Barium	Ba	56	137.34
Berkelium	Bk	97	(247.0702)
Beryllium	Be	4	9.0122
Bismuth	Bi	83	208.980
Boron	B	5	10.811
Bromine	Br	35	79.909
Cadmium	Cd	48	112.40
Calcium	Ca	20	40.08
Californium	Cf	98	(251)
Carbon	C	6	12.01115
Cerium	Ce	58	140.12
Cesium	Cs	55	132.905
Chlorine	Cl	17	35.453
Chromium	Cr	24	51.996
Cobalt	Co	27	58.9332
Copper	Cu	29	63.54
Curium	Cm	96	(247)
Dysprosium	Dy	66	162.50
Einsteinium	Es	99	(254.0881)
Erbium	Er	68	167.26
Europium	Eu	63	151.96
Fermium	Fm	100	(253)
Fluorine	F	9	18.9984
Francium	Fr	87	(211.996)
Gadolinium	Gd	64	157.25
Gallium	Ga	31	69.72
Germanium	Ge	32	72.59
Gold	Au	79	196.967
Hafnium	Hf	72	178.49
Helium	He	2	4.0026
Holmium	Ho	67	164.930
Hydrogen	H	1	1.00797
Indium	In	49	114.82
Iodine	I	53	126.9044
Iridium	Ir	77	192.2
Iron	Fe	26	55.847
Krypton	Kr	36	83.80
Lanthanum	La	57	138.91
Lawrencium	Lw	103	(257)
Lead	Pb	82	207.19
Lithium	Li	3	6.939
Lutetium	Lu	71	174.97
Magnesium	Mg	12	24.312
Manganese	Mn	25	54.9380
Mendelevium	Md	101	(256)
Mercury	Hg	80	200.59
Molybdenum	Mo	42	95.94
Neodymium	Nd	60	144.24
Neon	Ne	10	20.183
Neptunium	Np	93	(237.0480)
Nickel	Ni	28	58.71
Niobium	Nb	41	92.906
Nitrogen	N	7	14.0067
Nobelium	No	102	(255)
Osmium	Os	76	190.2
Oxygen	O	8	15.9994
Palladium	Pd	46	106.4
Phosphorus	P	15	30.9738
Platinum	Pt	78	195.09
Plutonium	Pu	94	(244)
Polonium	Po	84	(209)
Potassium	K	19	39.102
Praseodymium	Pr	59	140.907
Promethium	Pm	61	(145)
Protactinium	Pa	91	(231.0359)
Radium	Ra	88	(226.0254)
Radon	Rn	86	(222.0175)
Rhenium	Re	75	186.2
Rhodium	Rh	45	102.905
Rubidium	Rb	37	85.47
Ruthenium	Ru	44	101.07
Samarium	Sm	62	150.35
Scandium	Sc	21	44.956
Selenium	Se	34	78.96
Silicon	Si	14	28.086
Silver	Ag	47	107.870
Sodium	Na	11	22.9898
Strontium	Sr	38	87.62
Sulfur	S	16	32.064
Tantalum	Ta	73	180.948
Technetium	Tc	43	(98)
Tellurium	Te	52	127.60
Terbium	Tb	65	158.924
Thallium	Tl	81	204.37
Thorium	Th	90	232.038
Thulium	Tm	69	168.934
Tin	Sn	50	118.69
Titanium	Ti	22	47.90
Tungsten	W	74	183.85
Uranium	U	92	238.03
Vanadium	V	23	50.942
Xenon	Xe	54	131.30
Ytterbium	Yb	70	173.04
Yttrium	Y	39	88.905
Zinc	Zn	30	65.37
Zirconium	Zr	40	91.22

Note: The figures in parentheses are the atomic masses of the most stable isotopes of radioactive elements. The other masses are the average for all the isotopes together, as they occur in their natural abundances.

ANSWERS TO SHORT QUESTIONS

chapter 1

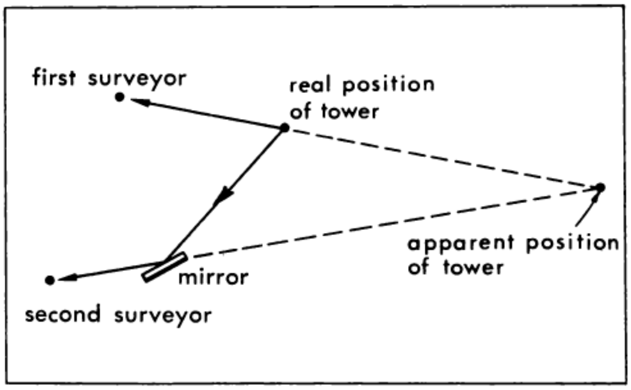
4. The stroboscope will appear to stand still, no matter how fast it is turned.
9. 10^5.
13. 300 m.
15. 9×10^{-2} mm.
16. They are the same to the nearest order of magnitude.

chapter 2

5. W is tripled also.
6. No matter what the value of k is, the graph is a straight line which passes through the origin of the coordinate system.
8. The total surface area is quadrupled.
9. 10^{12} maps.
12. 11 m^2.
27. (a) 27. (b) 27. (c) 9. (d) 9.

chapter 3

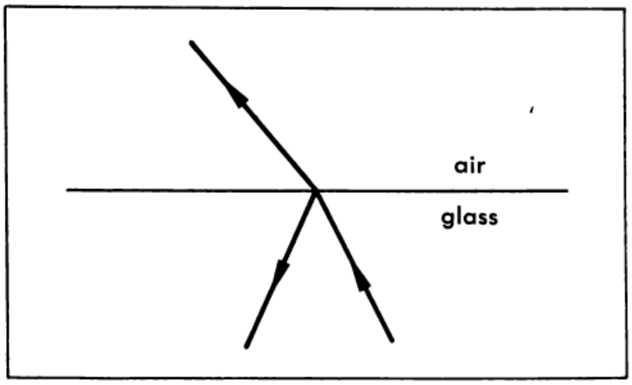
1. Firefly, flash bulb, and electric stove heating element.
2. Look at it in the dark.
4. Object A will appear reddish.
   Object B will appear very dark.
11. It is hotter and hence it emits more infrared radiation.
12. No; ultraviolet radiation cannot pass through glass.
13. When light is reflected or when it is refracted (bent) as it passes from one material into another.
18.



chapter 4

1. Angle of incidence = $\measuredangle 3$.
   Angle of reflection = $\measuredangle 2$.
2. An infinite number of rays, which together form a hollow cone whose axis is perpendicular to the mirror surface.
3. One.
8. Angle of incidence = $\measuredangle 2$.
   Angle of refraction = $\measuredangle 4$.
10. In the photo of Fig. 4-11.
12. On the left.
13. $r = \frac{3}{4}\, i$.
14. $n = \sqrt{3}/\sqrt{2} = 1.2$.
15. Yes, since the measured n is 1.3 to 1.4. The material is not necessarily water.
16. Diamond.

19.



23. $n = \frac{\sin \theta_A}{\sin \theta_B} < 1.$

28. 63°.

31. Yes; a prism will spread out the light into a spectrum so that you can see which colors have been filtered out.

chapter 5

1. The smaller Ping-Pong ball is deflected by small bumps and holes that would not affect a basketball.

3. The speed on the lower level is the same for all angles of incidence.

4. The speed of light should be greater in a refractive material than in a vacuum.

8. Isolate each one in a dark room where there are no reflections, and see whether each gives the same reading on a light meter placed the same distance from it.

9. At $r/\sqrt{2} = 0.71\ r$.

14. 2 meters.

21. The light is able to travel through the hole to the distant mirror and back through the hole before the disc has turned far enough to move the hole out of position.

29. The particle model predicts that the speed of light is greater in water than in air, while actual measurements show that it is less in water than in air.

chapter 6

1. The waves all travel through some medium, but the medium itself does not travel with the waves.

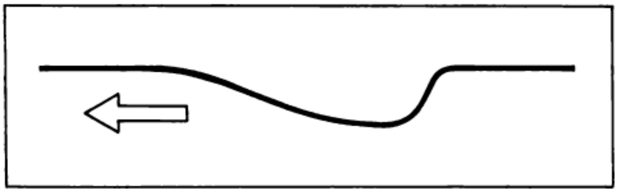
3. At the point of maximum displacement.

4. It would still move up first.

7. No.

8. 7 cm.

12.



16. The incident pulse.

17. 3. (The speeds are proportional to the distances the pulses move between successive pictures.)

22. The pulse is smaller in the bottom frames than in the top frames.

chapter 7

1. About 20 cm long.

2. $r' = 25°$.

3. (a) Incident pulse: (2); reflected pulse: (1). (b) $i = 70°$.

10. (a) 85 m. (b) It decreases.

11. (a) $T = 1/4$ sec. (b) $f = 4$ per sec.

12. Decrease the frequency.

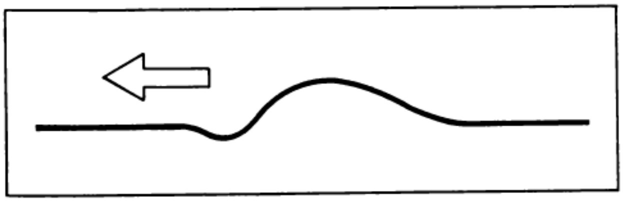
17. Near the fifth crest from the bottom.

18. $n \approx 1.8$.

20. No, not unless one frequency was a whole-number multiple of the other.

chapter 8

1.



2. A distance λ from the wall.

4. "Double crest": *A;* "double trough": *C;* nodal point: *B.*

7. No; along this line are double crests separated by double troughs.

10. The second nodal line to the left of the central line.

11. On which nodal line the point lies—that is, the value of *n*.

14. Yes; all λ such that $\lambda > 2d$.

21. $p = 0.36$.

22. Yes. In both cases the two sources dip simultaneously.

chapter 9

1. Try to find places where the boat does not bob up and down. Along the nodal lines the water does not move, while to either side of them the boat will bob up and down.

7. 0.9 meter.

9. $\Delta x = 0.68$ mm.

10. The nodes for red light are about 1.4 times as far apart as those for blue light.

13. All the sources producing the central maximum are in phase. Of the sources producing the partial reinforcement in Fig. 9-11, two thirds cancel each other, and the remaining one third are not in phase but are separated by phase delays of less than 1/2.

19. It would increase.

20. $\lambda = 4200$ Å.

21. No; the wavelength within each material depends on the index of refraction of the material.

22. The bars would be thinner and only half as far apart.

chapter 10

1. (a) $f = 24$ mm. (b) There is no focal length for a plane mirror.

2. The rays would be reflected back to the parabolic mirror, where they would then be reflected parallel to the axis.

3. Two rays are enough, because the others pass through almost the same spot.

6. (a) n_{liquid} is lower than n_{lens}. (b) n_{liquid} equals n_{lens}. (c) n_{liquid} is higher than n_{lens}.

13. Yes; whenever $S_o < f$.

33. The magnification is the same. Check the distance between the centers of the images.

chapter 11

1. (a) 120 km. (b) 20 km.

5. Graph (a).

10. No. The graph indicates that between $t = 2$ and $t = 3$ hours, the car is in two different places at the same time.

15. $v_{\text{ave}} = 0.48$ m/sec.

18. The velocity is zero at $t = 67$ sec, but the acceleration is never zero.

23. In the right-hand *straight* portion of the graph.

chapter 12

1. 2 meters (north), no matter in what order you walk the distances.

5. The first is a velocity, 30 km/hr west, while the second is a displacement, 30 km west.

7. The flag will hang limply.

14. (a) −86.6 km/hr along north and −50.0 km/hr along east. (b) −86.6, −50.

15. Zero.

18. Each velocity vector would have to be turned around to point in the opposite direction, but its tail would remain in the same place. Its magnitude would be unchanged.

20. No; it can be along any path.

21. If the motion is along a *curved* path, the *direction* of the velocity changes, requiring that there be an acceleration.

25. The swing will appear to be at rest.

26. It appears to move away at constant velocity.

chapter 13

1. Kinematics is the study of the motion of objects, without considering the cause of the motion; dynamics is the study of how forces cause motion.

2. No. Time was not involved in his reasoning.

4. The ball would orbit forever.

6. $v = 20$ cm/sec.

7. $v = 20.9$ cm/flash.

8. The force is constant.

12. Toward the right.

13. $a = 1.5$ m/sec^2.

16. (a) 3.0. (b) 3.0.

17. $a = 4$ m/sec^2.

18. $F = 2$ newtons.

19. $F = 15$ newtons.

25. Along the direction of the net force.

26. $a = 8$ m/sec^2 (to the right).

27. $\vec{\mathbf{F}} = 3.7$ newtons, to the right along the dotted line.

32. $\vec{\mathbf{F}} = 200$ newtons, in the direction opposite to that of your push.

chapter 14

1. $m_g = 5.00$ kg.
2. (a) Weight on moon: 1.1×10^2 newtons. Weight on earth: 6.9×10^2 newtons. (b) Mass on earth: 70 kg. Mass on moon: 70 kg.
4. $g = 9.81$ newtons/kg.
6. 4.9 meters long.
13. The horizontal component of displacement of the ball is constant over equal time intervals; on the other hand, the vertical component of displacement is not constant.
17. $F = 24$ newtons.
18. The speed does not change, but the direction of motion changes in the direction of the force.
19. $\vec{v} = 0.52$ m/sec, toward the left if the puck is moving clockwise, as in Fig. 14-11.
20. $\vec{a} = 0.64$ m/sec^2, toward the center of the circle.
26. $\vec{F} = -k\vec{x}$, for some positive constant k.
27. At the points where the circle intersects the y axis.
28. $m = 0.2$ kg.
29. $l = 0.25$ meter.
39. Slightly west of due north.

chapter 15

2. (a) 1.2. (b) About 6.
6. 4.75×10^{23} m^2.
8. (a) When it is closest to the sun. (b) When it is farthest from the sun.
9. 3.35×10^{18} m^3/sec^2.
10. $(8^3)^{1/2} = 23$ times longer.
13. 3.6×10^3 times greater.
14. (a) $K_e = 9.9 \times 10^{12}$ m^3/sec^2. (b) $m_e/m_s = 3.0 \times 10^{-6}$.
15. At a height equal to $(\sqrt{2} - 1) = 0.41$ times the earth's radius.

chapter 16

1. 1.5 newton-sec.
2. 32.0 kg-m/sec = 32.0 newton-sec.
3. (a) Four: the throw, the bounce, the catch, and the continuous impulse due to gravity.

   (b) The bounce. Here the ball experienced about twice the change in velocity that it did during the throw or catch, and the gravitational impulse was the smallest of all.
8. $\vec{F} = 500$ newtons, in the opposite direction to the momentum.
12. Zero. The net change in momentum is always in the direction of the total impulse.
13. Each velocity vector in (a) is multiplied by the mass of the body whose motion it represents, and a new scale for the units is chosen to represent these new vectors.
14. (Mass of faster)/(mass of slower) = 2/5.
21. The direction of motion of the center of mass after the collision is the same as it was before the collision: it is upward, at an angle of about a degree counterclockwise from the direction of motion of the ball entering from the bottom of the picture.
30. Most of the momentum is transferred to the earth (through the road) and a small amount to the air molecules surrounding the car.
32. No; the vector sum of the impulses due only to the forces of interaction cannot be zero, because the total impulse, which includes the momentum of the light, must equal zero.

chapter 17

4. Yes; no energy is transferred when a force acts on a moving body at right angles to its motion.
6. 200 joules.
7. The kinetic energy increases by a factor of four.
8. They will be equal, since they depend only on the force and the kinetic energy.
14. They repel each other during the interaction.
16. When the force between the colliding bodies depends only on the separation of the bodies.
17. On the way in, the putty ball experiences a repulsive force at small separation, which slows it down; after the putty gets to the wall, the force suddenly becomes attractive, so it is not the same as on the way in. Hence we know the force does not depend on distance alone.
18. Probably. Since kinetic energy is conserved, we may assume that in this collision the interaction force is a function only of separation.
19. Mass of $m_1 <$ mass of m_2.

chapter 18

1. In the middle picture, where the force is greatest
2. The slope of the graph is k.

9. Each would move in a straight line in the direction of the force acting on it.

17. 3.4×10^3 joules.

24. An additional E_K.

25. $GMm/\ (10r_e)$.

26. They are equal.

31. $F = 0.20$ newton.

chapter 19

1. (a) Wherever $\Delta\vec{v}$ points inward [in all of Fig. 19-1 (a) and in flashes 2-12, 14-27, and 30-41 of Fig. 19-1 (b), counting counterclockwise the successive magnet positions].

   (b) Wherever $\Delta\vec{v}$ points outward [in flashes 12-14 and 27-30 of Fig. 19-1(b)].

   (c) The magnetic force plus the component of the gravitational force perpendicular to the line of suspension.

2. Yes.

5. The same as it is when the object is at B, since $(\overline{OB})_\perp = (\overline{OA})_\perp = \vec{r}_\perp$.

6. Not unless $\vec{r}_\perp$ is the same for both. Note that the bodies do not have to be at the same position in order for $\vec{r}_\perp$ to be the same.

11. Yes; they will if their initial speeds are the same.

12. At the point of closest approach.

13. (a) The value of v_r is zero at the ends of the major axis of the ellipse.

    (b) At the end closer to the earth, $v_\perp$ is greater than at the end farthest from the earth.

14. Hint: Let $\Delta = \sqrt{b^2 - 4\,ac}$, $\Delta^2 = b^2 - 4\,ac$.

19. (a) Area projected on x-y plane $=$ $(10 \text{ cm}^2)\ (\cos 30°) = 8.7 \text{ cm}^2$.

    (b) Area projected on y-z plane $=$ $(10 \text{ cm}^2)\ (\sin 30°) = 5.0 \text{ cm}^2$.

20. (a) Into the plane of the paper, perpendicular to it.

    (b) Out of the plane of the paper, perpendicular to it.

24. $\omega = 3\pi$ radians/sec.

27. $v = 63$ cm/sec.

28. (a) $\omega = 0.20$ radian/sec.

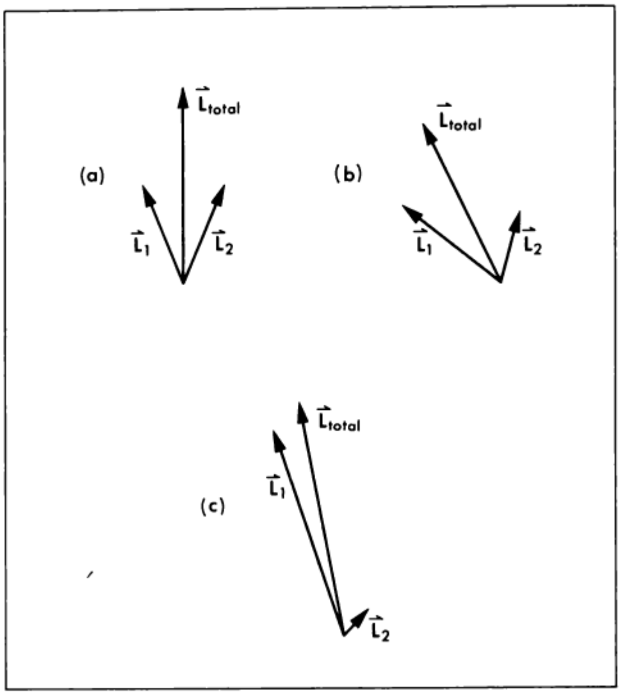
    (b) (i) $\omega = 0.20$ radian/sec.

    $$\text{(ii) } \omega = \frac{(3.0 \text{ m/sec})\ (\cos 30°)}{\dfrac{15\text{m}}{\cos 30°}} = 0.15 \text{ radian/sec.}$$

29. $I = \frac{1}{2}\ ml^2$.

30. $R/\ \sqrt{2}$.

31.



41. Yes.

42. The units are (magnet mass) (cm²)/(flash).

chapter 20

1. The impulse given to the wall is $2mv$, by conservation of momentum.

3. $P = F/A$.

5. The total kinetic energy of the center-of-mass motion is proportional to PV/N.

7. P increases by a factor of $373/298 = 1.25$.

8. E_K increases by a factor of $373/298 = 1.25$.

9. 45°C.

10. No.

14. Zero, if the container is not moving.

16. So that the kinetic energy of the masses can be ignored in comparison with the change in gravitational potential energy and the heat which the water acquires.

18. The internal energy does increase: mechanical energy supplied by the hammer is converted into internal energy, but heat does not flow in from the outside.

19. No. Energy is transferred through heat flow, not by doing mechanical work.

20. The total internal energy remains constant in this closed system.

21. 1×10^3 joules.

26. Gravitational potential energy is converted only

into heat energy, since the car does not speed up.

chapter 21

2. About 50 percent.
3. (a) About 50.
   (b) About 25.
4. HH, HT, TH, TT. The same side appears both times in 50 percent of the cases.
5. The probability is $1/2$.
6. (a) $1/6$
   (b) $2/36 = 1/18$.
9. (a) No.
   (b) No.
   (c) Yes.
10. (a) Specify which squares have white (or black) faces showing.
    (b) Specify how many white (or black) faces are showing.
12. Tilting 10,000 identical machines at the same time, or tilting the same machine 10,000 times successively, will get you the same statistical results. From Fig. 21-7 one can see that the probability for equal distribution is 0.25, so one would expect close to 2500 machines to have equal distribution.
13. (a) It is the probability that slot 1 contains 0, or 1, or 2, or . . . . N marbles.
    (b) Since slot 1 must contain one of these, the probability must be 1.
22. $3/4$.

chapter 22

1. By specifying the number of marbles in each slot.
2. "Macroscopic" quantities are (a) and (d); "microscopic" quantities are (b) and (c).
3. It would be greater than T.
5. The transfer of internal energy between the gas and its container is not instantaneous.
6. Zero surface area.
7. α' would be nearly zero. The only damping would be due to mechanical friction.
12. (a) $\left(\frac{V_0 - V}{V_0}\right)^N$
    (b) $V < V_0/2$.
13. $e^{-\Delta S}$.
15. The vapor.
16. $\Delta S = \pm 1.8 \times 10^{24}$.
18. $\Delta S = \frac{(l - l\cos\theta)\, mg}{kT} \approx \frac{mgl\theta^2}{2kT}$ for θ (in radians) much less than unity.

chapter 23

1. Negative.
2. No. The removal of any number of negative charges from a neutral object gives it a net positive charge.
3. (a) Charge flows through the metal rod onto your hand and through your body, so no charge can accumulate on the rod. (b) Insulate it from everything except what you are charging it with.
6. (a) Nothing. (b) Nothing. (c) Both electroscopes will discharge completely.
7. Positive.
8. Negative.
13. They will fall.
14. In a good vacuum there are hardly any charged ions which can remove charge from the leaves.
16. The electroscope leaves would fall.

chapter 24

1. $F = 1.0 \times 10^{-5}$ newton.
8. The net force would change in both magnitude and direction.
9. 24-8(a) and 24-9(a)
   24-8(b) and 24-9(b)
   24-8(c) and 24-9(c)
   24-8(d) has no analog in 24-9.
10. No. If they did, at the point they crossed the *net* force on a movable charge would have to be in two directions at once, which is impossible.
11. Only if the line is straight and the particle is moving along the line. If the line is curved, the trajectory of the particle will, in general, be different from the direction of the force on the particle.
17. $F = 1.4 \times 10^{-15}$ newton.
23. $F = 2.3 \times 10^{-16}$ newton.

chapter 25

1. 5.6×10^{-17} joule.
3. 7.64×10^{-6} sec.
6. $I = 2 \times 10^{16}$ elementary charges/second.

7. $I = 3.6 \times 10^{16}$ elementary charges/second.
11. 4.50×10^{22} atoms of copper.
13. 2.9×10^{-17} joule.
14. $m = 2.7 \times 10^{-26}$ kg.
20. $F = 4.2 \times 10^{-14}$ newton.
21. 1.7 joules/sec = 1.7 watts.
22. The potential energy at a distance r is $U = kq$ (1 elem. ch.)/r. All this will be transferred to the moving charge as kinetic energy at infinity.
23. $\vec{E} = 9 \times 10^{-18}$ newton/elem. ch., pointing toward the object. $V = 5 \times 10^{-18}$ joule/elem. ch.
30. $\varepsilon = 1.9 \times 10^{-18}$ joule/elem. ch. = 12 volts.
31. $I = 5.00$ amp.
32. 14 ev.

chapter 26

4. The large number of particles that come through the foil undeflected, compared with the small number that are deflected, tells us that the chance for one collision is very small; hence, the chance that one alpha particle will have two collisions is much, much smaller.
5. An atom must be mostly empty space, with something in it which can exert a strong force on an alpha particle.
6. It is the angle between the original and final directions of motion of a scattered particle.
10. We cannot measure b directly, because we do not know where each nucleus is, nor do we know the exact trajectories of the incident alpha particles.
11. N_θ is proportional to the square of the aiming error.
12. It is a Coulomb field; that is, it varies as $1/r^2$.
13. Through angles between 90° and 180°.
19. About 6 times as close.
20. $r_0 = 7.0 \times 10^{-15}$ meter.
21. The law of conservation of energy.

chapter 27

1. The electric field is also in the direction of the wire.
2. $E = 3.0$ volts/meter = 3.0 newtons/coulomb.
3. The electric field in the first wire will be the same as before, since the difference in potential from one end of the wire to the other still equals the EMF of the battery, no matter what else has been connected in parallel.
4. $V = 18$ volts.
9. If it removes much charge it will change the potential difference it is supposed to measure.
10. The current in the voltmeter must be small compared with the current in the circuit.
14. 45 electron volts = 7.2×10^{-18} joule.
15. If the negative terminal is connected to the right-hand plate, as in Fig. 27-8, the electron will give up 30 electron volts of its kinetic energy in order to do work against the repulsive force. This energy is transferred to the 30-volt battery. When the electron hits the plate, it will give up the other 15 electron volts as heat.
16. $R = 4.4 \times 10^4$ ohm. [The potential difference across the resistor is (90 - 37) volts.]
22. $V = 4.9$ volts.

chapter 28

1. The filings would try to stand up at right angles to the paper.
2. No; they are all closed loops, of various shapes.
3. Clockwise, in both cases.
4. The magnitude of the resultant field is $\sqrt{2}$ times the magnitude of the field from one.
5. The field is proportional to the magnitude of the current.
9. Parallel to the field, in either the same or opposite direction as the field.
10. Downward.
11. $f = 2.1$ newtons.
16. The force is directly proportional to the angle of rotation.
17. Switch the connections to the battery terminals, or reverse the direction of the magnetic field—but not both.
18. The radius increases as the square root of the accelerating potential.
19. The velocity of the charge, the sign and magnitude of the charge, and the component of magnetic field perpendicular to the velocity.
23. 360 electron volts = 5.76×10^{-17} joule.
30. $2\pi KI$.
31. Zero.
32. $B = 2\pi \times 10^{-3}$ newton/amp-m $= 6 \times 10^{-3}$ newton/amp-m.

chapter 29

1. The positive charges will be at Q, the negative charges at P.

2. (a) The force on positive charges is directed outward from the page; the force on negative charges is directed into the page. (b) No; charges are not free to move in that direction.
4. There would be no current as long as the entire loop was in the region of uniform field.
5. (a) When the plane of the loop is parallel to the magnetic field. (b) When the plane of the loop is perpendicular to the magnetic field.
6. The flux through the loop will not change.
7. $\phi_B = 1 \times 10^{-2}$ newton-m/amp.
9. Yes. The force you must exert on the loop is proportional to the current, which would be reduced if the electrical resistance of the loop were increased.
10. At high speed, since the pulling force would have to be greater.
11. $\varepsilon = 1 \times 10^{-4}$ volt.
12. $\varepsilon = 0.5$ volt.
19. No; the magnetic force on the charges in the moving loop, which is proportional to the speed of the loop, would increase until it exactly canceled the applied force, and the loop would then move with constant velocity.
24. No; the lines are closed loops, since there are no free charges on which they can begin or end.
26. The rate at which charge is flowing onto either plate, which in turn determines the rate of change of electric flux through the region between the plates.
27. (a) Yes; with a constant current. (b) No; it is impossible to change the magnetic flux within a region at a constant rate indefinitely.
28. (a) Negative charges are on the top plate, positive charges on the bottom plate. (b) The electric field points upward.
30. (a) 3×10^8 m/sec. (b) 3×10^8 m/sec. (c) 3×10^8 m/sec.

chapter 30

1. t_{AB} approaches $d/2u$.
   t_{BA} approaches infinity.
6. It should propagate at an angle of 60° to the bank.
7. $\sqrt{1+x^2}\left(\frac{1}{1-x}\right) \approx \left(1+\frac{x^2}{2}\right)(1+x+x^2) \approx 1+x+\frac{3}{2}x^2.$
13. At the right.
17. (a) At A.
    (b) Point A' is the point from which the light appears to come.
18. Points S_1 and S_2 are the slits in the double slit. Point A is the single slit. Point A' is to the right of the single slit.
19. (a) To the left through an angle $\theta = v/c$.
    (b) To the right through an angle $\theta = v/c$.
    (c) No shift.
    (d) To the left through an angle $\theta = (v/c)(n^2 - 1)$.
20. $\frac{d}{\lambda}\sin\alpha$.
21. 0.27, according to simple addition of velocities.
23. $w = \dfrac{u-v}{1-\dfrac{uv}{c^2}}$.
28. No noticeable phase shift. Indeed, he could detect no phase shift.
29. $T + d/c$.
30. $+d/c$.
31. (a) $vt' = vd/c$.
    (b) dv^2/c^2, since $t = (d/c)(1 + v/c)$ at this instant for all observers in the unprimed system. The difference is second order in v/c.
32. The clock at A' in the primed system reads the same as all the unprimed clocks.
33. Yes. The factor $1/\sqrt{1-v^2/c^2}$ in numerator and denominator cancels.
37. In neither.

chapter 31

1. The time at which a first jump occurred would always be later for weak light than for strong light.
5. (a) No. (b) Yes.
9. The points corresponding to frequencies of 6.8 and 7.4×10^{14} cycles/sec.
10. 2.4 ev.

chapter 32

1. $v = 2.5 \times 10^8$ m/sec.
2. Relativistically, $E_K = 0.250\ mc^2$, while classically, $E_K = 0.180\ mc^2$. The relativistic energy is almost 40 percent greater than the classical energy.
3. Multiply the numbers on the energy scale by the ratio of the proton mass to the electron mass.
9. Relativistically, $p = 0.750\ mc$, while classically, $p = 0.600\ mc$. The relativistic momentum is 25 percent greater than the classical value.
10. 1.88×10^{21} Mev/c.

11. $p \sim 10^3$ Mev/c, since the rest energy is negligible compared to the kinetic energy and the speed is very close to c.

27. $\nu = 1.3 \times 10^{20}$ cycles/sec.
$\lambda = 0.024$ Å.

41. (a) Toward the blue frequencies.
(b) Upward.

chapter 33

1. By Kepler's third law, R^3/T^2 is constant, so the ratio of the frequencies is $\sqrt{8} = 2.8$.
2. 4.2 ev.
6. 6.67 ev, 4.86 ev, 1.81 ev.
7. 1.81 ev.
12. The atom absorbs 4.86 ev from the electron. It absorbs no energy from the photon. [See Fig. 33-4(b) and 33-7(c).]
16. $-$ 10.4 ev.
17. The γ line of the Lyman series.
18. (a) 13.5 ev. (b) $-$ 0.136 ev.
19. (a) Any energy from 10.2 ev up to 12.1 ev. (b) 12.1 ev, 10.2 ev.

chapter 34

1. $\lambda = 0.73$ angstrom.
2. $2d \sin \theta = 2\lambda$.
12. No; since the only restriction is that $2d$ equal a whole number of wavelengths, there are an infinite number of energy values which the particle could have; d determines the lowest energy level and also the energy values of the higher levels which can occur.
13. 10^{-17} joule.

chapter 35

1. Although kinetic energy is positive, while total energy is negative, they are of comparable magnitude. For hydrogen:

$$\frac{E_K}{mc^2} \sim \frac{R}{mc^2} \sim \frac{10 \text{ ev}}{10^6 \text{ ev}} = 10^{-5}$$

Hence,

$$E_K = \frac{1}{2} mv^2 \text{ is valid.}$$

6. The lines of wavelength 303Å, 256Å, and 243Å, respectively.
7. One half of a Bohr radius, or 0.26×10^{-10} m.
8. In the infrared or possibly in the visible, certainly not in ultraviolet, since $Z_{\text{eff}} \approx 1$ only for large values of n.
13. No; the most energetic photon given off in such a transition would have an energy of only $\Delta E = E_\infty - E_2 = 2.25$ rydbergs.
19. Such points would correspond to a value of Z_{eff} greater than Z, the total charge of the nucleus.
20. 8.8×10^4 ev. Note that this is not equal to RZ^2 for $Z = 82$.
26. $n = 3$.
27. $n = 4$.
30. Yes; lead, with 82 electrons in each neutral atom, must have electrons in shells beyond the first and second, and these electrons could be knocked out by photons of longer wavelengths.
32. For (3, 1, 1), $E = 11(h^2/8\ md^2)$
For (2, 2, 2), $E = 12(h^2/8\ md^2)$
33. A total of six electrons can have the same energy as a (3, 1, 1) electron. Only two can have the energy of a (2, 2, 2) electron.
34. Distance ratio = 4.32; $E = 2.22$ Mev.
37. In inelastic proton scattering, the energy lost by the proton corresponds to the energy gained by the nucleus in transition from the ground state to each of the various excited states. But in slow neutron capture, the nucleus is produced in an excited state and subsequently makes transitions through a large number of intermediate states to the ground state, in the process emitting numerous gamma rays whose energies correspond to the energy differences between all these various levels.
38. The 0.09 Mev is the recoil energy of the Mn^{55} nucleus.

chapter 36

1. About 1.5 ev.
2. At total energy equal to 2.1 Mev.
4. In the center-of-mass frame of the positron and electron, the total momentum before annihilation is zero. Yet one photon would have a momentum of $2mc$, so momentum would not be conserved.

6. Conservation of charge and conservation of energy.

9. No, they are in order of increasing binding energy per particle. The order should be: He^4, O^{16}, Ni^{60}, Mo^{96}.

10. $m = Zm_p + (A - Z)m_n - AB/c^2$, where c is the speed of light. For convenience in computation, multiply through by c^2 and express each rest energy mc^2 in the same units as B.

11. Total binding energy of Li^6 ($5.3 \times 6 = 31.8$ Mev) exceeds the combined total binding energy of He^4 and H^2 ($7.1 \times 4 + 1.1 \times 2 = 30.6$ Mev).

15. Rn^{218}, with a shorter half-life, has a greater probability of decaying.

16. The kinetic energy inside the nucleus is much greater, since the potential energy is negative there but the total energy is the same.

23. $\tau \Delta E \approx 10^{-15}$ ev-sec; $h = 4 \times 10^{-15}$ ev-sec.

INDEX

A

Aberrations, chromatic and spherical, 160-162
Absolute value, 200
Absorption of light, 44–45
  in particle model, 81–82
Absorption spectra, 557, 631–634, 656–659
Acceleration, 185–189
  and force applied, 226–228, 231
  average, 187
  centripetal, 243
  in free fall, 236–238, 237 (Table)
  instantaneous, 186
  of frames of reference, 251–253
  of projectiles, 206–207, 238–239
  vector, 204–212
Accuracy, 13–14
Adams, John Couch, 272
Adiabatic gas law, 415
Adiabatic processes, 408, 414–415
Age paradox, 583–584
Aiming error, 488
Alkali metals, atomic spectra of, 632
Alpha Centauri, 30
Alpha decay, 677–680
  lifetime of, 678, 686–687
  wave model of, 683–686
Alpha particles, 534-535
  charge of, 456
  closed chamber tracks of, 431, 485
  mass of, 535
  trajectories in nuclear fields, 344–345, 485–494, 677–679
Alphonso X, 260
Ammeter, 478
  moving coil, 523
Ampère, André Marie, 526
Ampère (unit), 478
  absolute, 529
Ampère's circuital law, 526
Amplification, by laser, 131
  in measurement, 8–9
Amplitude of simple pendulum 250
Anderson, C. D., 613
Angle of incidence, 55, 58, 102
  critical, 67
Angle of reflection, 55, 102
Angle of refraction, 58, 106
Angle of rotation, 351
Angle of scattering, 488
Angstrom unit, 478
Angular momentum, 343
  and moment of inertia, 352–354
  and trajectories, 345–346
  conservation of, 348–350, 359, 674–675
  orbital, 356–359, 674–675
  rate of change of, 355–356
  spin, 357, 662, 674–675
  vector nature of, 347–348
Angular velocity, 351
Annihilation, 613–614
Antennas, 556
Apollonius, 260
Arcs, 557
Aristarchus, 260
Aristotle and motion, 219, 231, 268
Arrangement, 389
Astrology, 258
Astronomical unit, 265
Atom, as light source, 129–131
  ionization energy of, 478
  Rutherford's planetary model of, 467, 485–486, 625
  stability of, 624–625, 647
Atomic clocks, 632
Atomic dimensions, 7, 478, 486, 660–661
Atomic energy levels, 627
  and matter waves, 644–647
  and spectra, 628–635
  of hydrogen, 634, 652
  of hydrogenlike atoms, 650–656
Atomic mass unit, 533
Atomic masses (Table), 533
Atomic number, 493, 657
Atomic spectra, 628–634, 652–660
  See also Spectral lines
Avalanche, 505
Avogadro, Amedeo, 369
Avogadro's number, 394

B

Bacon, Roger, 376
Balance, 11–12
  electric, 452–454
Balmer, J. J., 634, 652
Balmer series, 634
Barrier penetration, 679–685
Batteries, 428
  EMF of, 472–473, 477, 498
  in circuits, 498–500
  in series, 448
Beams of light, 49
Bertozzi, William, 605
Beryllium decay, 686
Betatron, 549
Binding, chemical, 663-664
Binding energy, atomic, 478, 655–656
  chemical, 662
  nuclear, 665, 675–677
  satellite, 333–334
Binocular vision, 51
Biot, Jean Baptiste, 524
Biprism, 640, 642
Black body, 596
"Black boxes," 14–15
Blue shift, 583, 618
Bohr, Niels, 652
Bohr radius, 653
Boiling, 374
Boltzmann's constant, 369
Bone strength, 36–37
Boyle, Robert, 376
Boyle's law, 376
Brahe, Tycho, 262–263

C

Calculus, 273
Calibration, 15
Calorie (unit), 378–379
Camera, 155
Candle (unit), 81
Cathode, 435
Cathode-ray oscilloscope, 435
Cavendish, Henry, 274–438
Cells, electrochemical, 477
  electrolytic, 467–470
  photoelectric, 46, 592–593
Center of mass, 287, 356–360
  motion of, 286–290, 349, 360, 374
Center-of-mass energy, 309
Central force field, 344–346
Centrifugal force, 252
Centripetal acceleration, 243
Centripetal force, 243–244, 252
  on a planet, 269

Cesium, energy levels of, 627
Chadwick, Sir James C., 493, 657
Charge, 439–441
  accelerated, 556
  conservation of, 455, 613
  electron, 456
  elementary, 445, 450–452, 470–471
  in conductors, 497–502
  induced, 425
  ionic, 469–470
  nuclear, 492–493
  unit of, 478–479
Charged particles, in a magnetic field, 529–531
  in a nuclear field, 486–489
Chemical binding, 663–664
Chemical properties, 660, 662
  and inertial mass, 227
Chemical reactions, 227, 477, 616
Chromatic aberration, 161
Circuit elements, 507–509
Circular motion, 211, 241–244, 248
Circular pulses, 100–102
Circulation, 525–526
  for electric fields, 550–552
Clocks, 4–6
  moving, 578–582
  synchronization of, 563–564, 575–577
Closest approach, 344–346
  and scattering, 493, 679
Cloud chamber, 431
Collisions, 76, 281–293
  and angular momentum, 348–350
  and energy transfer, 305–307
  and entropy changes, 412–413
  elastic, 309–310, 324–327, 395
  electron-positron, 612–614
  in a gas, 395–396, 505
  inelastic, 315, 325, 375, 397–399, 412
  in metals, 506
  in proton scattering, 666
  of photons and electrons, 611–612
  of subatomic particles, 312–313, 611–614
  temperature rise in, 375
Color, 44–45, 68–70, 69 (Table)
  and chromatic aberration, 161–162
  and wavelength of light, 108, 131–133, 132 (Table)
  in interference patterns, 132, 140
  in thin films, 137–140
  of the sky, 397
Comet, 488
Commutator, 524
Compass, 514–515
Complex nuclei, 666–668
Compression, 320–324
  of gases, 404–408
Compton, A. H., 611
Compton scattering, 611–612
Condensation, 374
Conduction, electrical, 422–423, 432, 497–502
  heat, 397–399, 413–414
Conductivity, of gases, 429–431, 504
  of metals, 433–434, 506
  of solutions, 432
Conductors, 497–502
  current-voltage relation in, 506
Conic sections, 488
Conservation of charge, 455, 613
Conservation of energy, 309-310, 329–335
  and energy states, 672–674
  and internal energy, 375–376, 379–381
  and matter waves, 644
  in Compton scattering, 611
  in electric circuits, 509
  in positron-electron annihilation, 614–615
  in the photoelectric effect, 596
Conservation of mass, and energy, 615–617
  in chemical reactions, 227
Conservation of momentum, 284–286, 290–292, 311
  angular, 348–350, 359
  in nuclear reactions, 674–675
  of photons, 611–612
Conservation of nucleons, 675
Coordinate transformation, 578, 582
Copernicus, Nicolaus, 213, 260–262
Cornu, M. A., 83
Correspondence principle, 652
Cosmic rays, 431, 580, 600, 616
Cosmological theories, 381
Coulomb, Charles A., 438
Coulomb (Unit), 478
Coulomb's law, 428, 440, 446
  and alpha-particle scattering, 486–492, 677–679
  and electric potential, 474–476
  constant in, 455
Crests, 100–101
Cross, W. G., 611
Cross product, 348
Crown glass, index of refraction, 69 (Table)
Crystal gratings, 558, 639–640
Crystal lattice, 398, 558, 600
Current, electric, *See* Electric current
Current balance, 528
Current-voltage curves, 504–509
Cutoff potential, 594
Cycloid, 213

D

Damping, 96, 403
Damping coefficient, 403–404, 407 (Table)
De Broglie, Louis, 638
De Broglie wavelength, 638–644, 641 (Table)
Delta notation, 179
Density, 25
  fluctuations of, 397
  of gases, 369 (Table)
Descartes, René, 62, 70
Deuterium, 614 (Table)
Deuteron, and hydrogen atom, 665 (Table)
  formation of, 616–617
  quantum description of, 664–665
Diffraction, and image resolution, 162–164
  and photons, 598
  double-slit, 129, 137
  electron, 640–642
  in microscopes, 164
  in telescopes, 148, 162–164
  neutron, 639–640
  of light, 132–136
  of radiation, 557
  of waves, 108–110
  single-slit, 134–136
Diffuse reflection, 54–55, 76
Diffusion, 415–416
Diffusion chamber, 431
Dilation of time, 578–582
Diode, 434
Direct proportion, 25–26
Dispersion, of light, 68–70
  of waves, 107–108
  See also Color
Displacement, 170, 173–175
  vector calculations of, 196–204
Displacement vector, 196
Distance, measurement of, 6–11, 7 (Table), 30
Distance of closest approach, 344–346, 493, 679
Distribution, 385, 389–394
  of scattering, 489–492
Doppler effect, 583, 618, 687
Dumbbell, 354
Dynamics, 213–214
  relativistic, 604–609

E

Earth, gravitational field of, 235–236, 236 (Table), 250, 328–332
  magnetic field of, 514
  mass of, 12, 273
  rotation of, 252–253, 260–262
Earth satellites, *See* Satellites
Eccentric circle system, 264
Eclipses, 48
Eddies, 214
Einstein, Albert, 141, 229, 275, 572, 575, 594, 596, 619
Einstein-Planck relation, 596
Elastic collisions, 76, 309–310, 324, 327
  and gas expansion, 395–396
  electron-photon, 611–612
Eldridge, John A., 630
Electric current, 429, 467–471
  and magnetic field, 515–522
  and potential difference, 504–509
  equivalent, 552–553
  induced, 541–546

  in photoelectric effect, 593–594
  saturation, 504
  unit of, 478
Electric field, 441–445, 474–476
  between charged plates, 445, 449
  in conductors, 499
  induced, 548–550
  of a nucleus, 486–488, 492
Electric flux, 552
Electric potential, 475–476
  measurement of, 502
  unit of, 477–478
Electrical conductivity, 429–432
Electrodes, 432
Electrolysis, 432, 467–471
Electromagnetic radiation, 553–558
  and photons, 599–601
Electromagnetic spectrum, 557
Electromagnetic waves, 553–558
  and photons, 599–601
  spectrum of, 557
  speed of, 555
Electrometer, 427–428, 435, 498, 502
Electromotive force, 473
Electron, 433–434
  charge of, 456
  mass of, 466, 471
  Pauli principle for, 662
Electron diffraction, 640–642
Electron emission, photoelectric, 592–597
  thermionic, 434, 505–506
Electron gas, 506
Electron gun, 435, 503
Electron shells, 661
Electron volt, 477–478
Electron-positron annihilation, 613–614
Electronic clock, 6
Electroplating, 432
Electroscope, 423–424, 427–428
Electrostatic induction, 425–426
Elements, chemical, 493, 533 (Table)
  ionization energies of, 659
  periodic table of, 660
Ellipse, 345, 488
E M F, 472–473, 477, 498
  induced, 546–550
Emission Spectra, 557, 628–631
Energy, and mass, 615–617
  binding, *See* Binding energy
  center of mass, 309
  conservation of, *See* Conservation of energy
  electric potential, 475–478
  escape, 332–334
  ionization, *See* Ionization energy
  internal, *See* Internal energy
  kinetic, *See* Kinetic energy
  mechanical, 334, 376–381
  potential, *See* Potential energy
  quantum, 596
  rest, 607, 615
  total, 329–330, 332, 379–381, 608, 615
Energy levels, 616, 627
  and matter waves, 644–647
  and spectra, 628–635
  effect of leakage on, 688–690
  nuclear, 666–668
  of hydrogen, 634, 652
  of hydrogenlike atoms, 650–656
  of the deuteron, 664–665 (Table)
Energy spread, 685–688
Energy transfer, 300–303, 305–315
  and temperature rise, 378–379
  by radiation, 380, 625
  in a battery, 472–473
  in elastic collisions, 309–310
  in electrolysis, 471
  in light reflection, 609–610
  in simple collision, 308–309
  kinetic and potential, 320–332
  mechanical and internal, 376–381
Entropy, 411
  and radiation, 417
  changes in, 412–415
  of a perfect gas, 415–416
  volume dependence of, 416
Equal-area law, 264, 341–342
Equipotential, 497
Equivalent current, 552–553
Errors in measurements, 12–14
Escape energy, 332–334
Escape velocity, 333
Ether, 141, 562, 566–571
Eudoxus, 259
Excitation energy, 627, 647
  and barrier penetration, 684
Explosions, 282, 286
Exponential functions, 31–34
Extrapolation, 23
Eyepiece, 158–159
Eyes, 46, 50–51, 155–157, 598

F

Factorial, 391
Faraday, Michael, 470, 540, 546
Field, 236
  See also Electric field, Force field, Gravitational field, Magnetic field, Nuclear field, and Potential
Field line, 442–443, 516–517
Films, thin, 133, 137–140
Filters, 45
  See also Color
Fission, 615, 675–677, 683–684
Fixed stars, 258, 261, 562
Fizeau, Armand, 48, 83, 572
Flash photography, 6
Fluctuation scattering, 397
Fluorescence, 43
Flux, electric, 552
  magnetic, 543–546
Focal length, 147, 151
Focal plane, 148
Focus, 144, 151
Focusing, in the eye, 155–157
  of optical instruments, 155
Foot-candle, 81
Force (electrical), 420–421, 439–441
  and potential, 474–476
  between elementary charges, 455
  nuclear, 665–666, 677–679
  on a moving charge, 471
Force (mechanical), 219–231
  and Newton's law of motion, 228
  and principle of inertia, 220, 290
  centrifugal, 252
  centripetal, 243–244, 252
  fictitious, 251–252
  gravitational, 235–236, 270–271
  linear restoring, 248–250, 323
  Newtonian, 291
  of interaction, 291–293
  vector calculation of, 229–231
Force field, 325
  central, 344
  inverse square, 345, 486–489, 492
  See also Electric field, Gravitational field, and Nuclear field
Foucault, Jean B. L., 83, 107, 252–253
Fractional error, 13–14
Frame of reference, 203, 212–213, 250
  accelerated, 251–253
  and planetary motion, 268
  and speed of light, 571–572
  experimental, 250–251
  moving, 574–582
Franck, James, 626
Franck and Hertz experiment, 626
Franklin, Benjamin, 421
Free expansion of gases, 394–396, 409–410
  and diffusion, 415–416
Free fall, 205, 236–238, 328–330
Freezing, 416–417
French, A. P., 610
Frequency, 104
  in Doppler effect, 583, 618
  of electromagnetic radiation, 556–557
Frequency spread, 685–687, 690
Frequency standard, atomic, 632
Friction, 220–221, 231
  and free fall, 238
  and kinetic energy loss, 314–315
Frisch, D. H., 580
Functions, exponential, 30–34
  linear, 25
  power, 26–29, 34–35
Fusion reaction, 685

G

Galileo, 36, 48, 376
Galileo's law of inertia, 220, 228, 343
  and moving frame of reference, 251
Galle, Johann, 273
Gamma rays, 557–558, 600, 614
  energy change of, 618
  See also Photon
Gas discharge, 504
Gas laws, 369, 414–415

Gas oscillator, 404–408
  entropy of, 412
Gas pressure, 371–372
Gases, conductivity of, 429–431
  densities of, 369 (Table)
  diffusion of, 415–416
  electric discharges in, 504–505
  free expansion of, 394–396, 409–410
  ideal, 369, 414–416
  ionization of, 430
  isothermal expansion of, 409–411
  kinetic theory of, 373–374
  molecular motion in, 314, 369–372
Gauss, Carl Friedrich, 522
Gauss (unit), 522
Geiger, H., 456, 491, 534
Geiger counter, 505
Geocentric system, 258–260, 262, 268
Geometrical solids, 263
Geometry, 13
Gilbert, William, 515
Glass, index of refraction of, 69 (Table)
Graphs, 20–23
Gratings, crystal, 558, 639–640
Gravitation, experimental test of, 274–275
  law of, 271
  universal, 268–274
Gravitational constant, 271, 273–274
Gravitational field, 235–236, 270–271
  and photon mass, 618
  of the moon, 247
  potential energy of, 328–332
  red shift of, 618–619
  work in, 332
Gravitational field strength, 236 (Table)
  and universal gravitation, 270
  measurement of, 250, 273
Gravitational mass, 227–228
  and theory of relativity, 275
Ground, electrical, 424

H

Hale telescope, 149
Heat, as internal energy, 376
  latent, 374
  frictional, 315
Heat bath, 409, 412
Heat capacity, 406, 414
Heat conduction, 397–399, 413–414
Heat engine, 379
Heat flow, 377–379
  in gas oscillator, 406–407
  in isothermal expansion, 410–411
Heat radiation, 47, 380, 417, 557
Heating by light absorption, 81–82
Heliocentric system, 267–268
Helium atom, 653–654
Helium isotopes, 614
Henry, Joseph, 542
Heracleides, 260
Herschel, Sir William, 272
Hertz, Gustav, 626
Hertz, Heinrich R., 555
Hipparchus, 260
Hirn, Gustave A., 379
Hooke, Robert, 270, 376
Horsepower, 331
Hull, G. F., 81, 609
Huygens, Christian, 107, 134, 270, 286
Hydrogen atom, energy levels of, 634–635, 647, 650–653
  formation of, 616
  Rutherford's model of, 486, 625, 635
Hydrogen ions, mass of, 461–465
Hydrogen isotopes, 614 (Table)
Hydrogenlike atom, 650–656

I

Ideal gas law, 369, 414
Illumination, 79–81
Illusions, 2, 58
Images, 45, 54–57, 103
  distortions in, 160–164
  formed by lenses, 151–156
  in telescopes, 147–148
  real, 152, 154–156
  resolution of, 163
  virtual, 56, 154–156
Impulse, 279
Impulsive force, 280
Incandescence, 43
Incidence, angle of, 55, 58, 67, 102
Inclined plane, 220
Index of refraction, 61–70, 63 (Table), 69 (Table)
  absolute, 64
  and chromatic aberration, 161–162
  and color, 69, 108
  and convergence of light, 149–150
  and particle model, 78
  of waves, 106, 108
  relative, 66
  reversibility of, 64–65
Induction, electromagnetic, 541–553
  electrostatic, 425–426
Inelastic collisions, 315, 325, 375
  and heat conduction, 397–399
  entropy changes in, 412–413
  in proton scattering, 666–667
Inertia, 220–221, 343
  and center of mass, 290
  moment of, 354
Inertial frame of reference, 252
Inertial mass, 225–228
  and free fall, 236
  and theory of relativity, 275
Infrared radiation, 47, 557
Insulators, 423
Intensity of light, 29–30
  and particle model, 79–81
  and photoelectric effect, 593
  in diffraction patterns, 136
Interactions, 281–293
  frictional, 314–315
  See also Collisions
Interference, and Michelson-Morley experiment, 566
  and photons, 598
  in a moving medium, 567–571
  in thin films, 137–140
  of light, 127–141
  of matter waves, 639–644
  of radiation, 557
  of water waves, 115–124
  on a spring, 114–115
Internal energy, 309, 374
  and conservation of energy, 375–376, 379–381
  and temperature, 375–378
  atomic, 627–628
Interpolation, 20–23
Inverse-square force field, 345, 440
  in heavy atoms, 658
  of a nucleus, 486–489, 492, 654, 665
Inverse-square relation, 29–30
  for electric charges, 438–439
  for gravitation, 270–271, 273
  for light intensity, 29, 79–81
Involute, 251
Ionization, 430
Ionization current, 504
Ionization energy, 478, 628, 657–660
  of helium, 654
  of hydrogen, 634–635
  of lithium, 655
  of the deuteron, 665
Ionized gases, 504–505
Ions, 430
  mass of, 532–533
Iraser, 557
Irreversibility, 385, 392, 413
Isothermal expansion, 409–411
Isothermal processes, 409
Isotopes, 533, 666
  and alpha decay, 677–678
  hydrogen, 614 (Table)
  masses of, 533 (Table)
  notation for, 675
Isotropism, 571

J

Joule, James, P., 303, 376
Joule (unit), 303
Joule experiment, 376–377
Jupiter, 48

K

Kelvin temperature, 368
Kepler, Johannes, 263
Kepler's laws, 263–266, 265 (Table), 270, 341–342

Kilogram, 228
Kinematics, 169, 213–214
  of planetary systems, 267
  relativistic, 582, 604–609
Kinetic energy, 304–305
  and frictional interactions, 314–315
  and momentum, 311–313
  and temperature, 373
  in elastic collisions, 309–310
  molecular, 315, 371–376
  relativistic, 606–608
  transfer of, 305–315
Kinetic theory of gases, 373–374

L

Laser, 131, 557
Latent heat, 374
Leakage, 681–685
  and energy spread, 685–686
  and sharpness of energy levels, 688–690
Lebedev, P. N., 81
Length contraction, 566–567, 582
Lenses, 149–156
  aberrations in, 160–162
  achromatic, 162
  and diffraction effect on images, 163–164
  in microscopes, 156–159
  in refracting telescopes, 159–160
  magnification by, 157–158, 160, 162
Lenz, H. F., 547
Leverrier, Urbain, 272
Lifetime, 580
  and energy spread, 685–686
  of alpha emitters, 677–678, 687
  of decaying states, 687
  of photon emission, 686–687
  of probability wave, 683
  of standing wave, 690
Light, absorption of, 44–45, 81–82
  and momentum, 292, 359
  convergence of, 145–146, 149–151
  diffraction of, *See* Diffraction
  dispersion of, 68–70, 107–108
  from a supernova, 292
  interference of, *See* Interference
  invisible, 47
  propagation of, 47–50
  reflection of, *See* Reflection
  refraction of, *See* Refraction
  speed of, *See* Speed of light
Light (models), electromagnetic radiation, 553–557
  particle model, 75–85, 590–591
  photon model, 589–601
  wave model, 100–110, 127–141, 562–571
  wave-particle, 644
Light emission, 292, 359, 557, 612
  and Rutherford atom, 625
  See also Atomic spectra and Spectral lines
Light intensity, 29–30, 79–81
  and photoelectric effect, 593
  in diffraction patterns, 136
Light pipe, 67
Light pressure, 81, 609–610
Limit, 183
Limitations in measurements, 5, 8–10, 13
Linear functions, 25–26
Lithium atom, 654–656
Lithium nucleus, dissociation of, 673–674
Logarithms, 31–32
Log-log plots, 34–35
Lorentz, H. A., 566
Lorentz contraction, 566–567, 582
Lorentz transformation, 578, 582
Luminosity, 43
Luther, Martin, 262
Lyman series, 634, 653

M

Macroscopic states, 402
Magnet, 514–518
Magnetic field, 514–518
  and electric current, 515–522
  circulation of, 525–526
  induced, 550–553
  near a wire, 524
  uniform, 526–528
Magnetic field strength, 519–522
Magnetic flux, 543–546
Magnification by lenses, 156–158, 160, 162
Mariotte, Edme, 286
Marsden, E., 491
Maser, 557
Mass, 11, 228
  and energy, 615–617
  center of, 286–290, 360
  gravitational, 227, 271
  inertial, 226, 236
  molecular, 369 (Table)
  of isotopes, 533
  of the earth, 12, 273
  photon, 617–619
  relativistic, 608
Mass elements, 352
Mass number, 533
Mass spectrograph, 532
Mass spectroscopy, 532
Matter, 11
  and energy, 616
Matter waves, 639–644
  and energy levels, 644–647
  leakage of, 681–685
Maxwell, James Clerk, 438, 548, 554
Mayer, Julius Robert, 376
Measurement, errors in, 12–14
  of mass, 11
  of space, 6–11, 7 (Table)
  of time, 3–6, 3 (Table)
Mechanical energy, and internal energy, 376–381
  total, 334
  See also Kinetic energy, Potential energy
Melting, 398, 412
Mercury, energy levels of, 627
  spectrum of, 630
Meson, 580, 600, 616
Metals, electrons in, 433–434, 506
Meter, moving-coil, 523
Michelson, Albert A., 83, 107, 564, 572
Michelson-Morley experiment, 564–566
Microbalance, 12
Micrometer, 8
Micro-microbalance, 448
Microscopes, 9, 156–159, 162–164
Microwaves, 557, 599, 632
Millikan, R. A., 450
Mirrors, 54–57, 103
  parabolic, 144–148, 153–154
  spherical, 146, 154, 160
Mole, 403
Molecular kinetic theory, 373–374
Molecular masses, 369 (Table)
Molecular radiation, 557, 687
Molecule, dissociation of, 672–674
  formation of, 663
Moment of inertia, 354
Momentum, 280, 281–284
  and kinetic energy, 311–313
  angular, *See* Angular momentum
  carried by light, 292, 359
  conservation of, *See* Conservation of momentum
  in radiation, 292
  linear, 344
  of wave-particle, 646
  relativistic, 608–609
Moon, motion of, 247, 269–270
Morley, Edward W., 564, 572
Motion, and force applied, 220–228
  and frames of reference, 212–213, 251–252
  and kinetic energy, 304–305
  circular, 211, 241–244, 248
  free fall, 236–238, 328–330
  in a central force field, 344–346
  in a magnetic field, 529–531
  in a nuclear field, 486–489
  linear, 171–189
  molecular, 315, 369–374
  Newton's laws of, 228, 281, 291
  of pulses on a spring, 88–96
  on a plane, 196
  perpetual, 221
  planetary, 258–271
  projectile, 206–207, 238–241, 330
  simple harmonic, 247–250, 323–324
  with constant acceleration, 187–189
Motors, 520, 523–524
Mount Palomar telescope, 149
Muon, 580–582
  and time dilation, 580–582

N

Natural logarithms, 31–32
Nebulae, 38
Neutron, charge of, 421
  discovery of, 313
  mass of, 614 (Table)
Neutron capture, 615, 664, 666
Neutron decay, 687
Neutron diffraction, 639–643
Newton, Sir Isaac, 70, 75, 246, 268–272, 376
Newton's law of motion, 218, 228–231
  and earth's rotation, 251–253
  in a moving frame of reference, 251–252
  in terms of momentum, 281
  vector, nature of, 231, 238
Newton's law of universal gravitation, 269–271
Newton's third law of motion, 291
Newton (unit), 228
Nichols, E. F., 81, 609
Nodal lines, 117–120
Nodes, 115
Nuclear dimensions, 10–11, 486, 494
Nuclear disintegration, 672–674
  and alpha decay, 677–680, 683–686
  and binding energy, 675–677
  and wave-particle model, 680–690
Nuclear field, 654, 658
  energy levels in, 650–653
  trajectories in, 344–345, 486–494
Nuclear fission, 615, 675–677
Nuclear force, 665–666
  and alpha decay, 677–679
Nuclear fusion, 685
Nuclear power, 615, 677
Nuclear reactions, 614–615
  and conservation laws, 672–675
Nucleus, 485
  charge of, 492–493, 657–659
  complex, 666–668
  energy levels of, 666–668
  radius of, 494

O

Objective lens, 158–159
Oersted, Hans Christian, 516
Ohm (unit), 506
Ohm's law, 506
Opacity, 45
Optical instruments, 144–164
  limitations of, 160–164
  resolution of, 163
Optical illusion, 2
Orbital angular momentum, 356–359, 674–675
Orbits, atomic, 653–661
  in inverse-square force fields, 488
  of artificial satellites, 245, 294, 333, 345–346
  planetary, 265 (Table), 269
Orders of magnitude, of distances, 7 (Table)
  of masses, 11 (Table)
  of time, 3 (Table)
Oscillations, and electromagnetic radiation, 556–557
  damped, 403
  in a crystal lattice, 398–399
  in standing waves, 644–647, 680–681
  on a spring, 247, 323–324
  simple-harmonic, 248–250
Oscillator, gas, 404–408
Oscilloscope, 435

P

Parabola, 146, 488
Paraboloid of revolution, 146
Parallel circuits, 508
Particle in a "box," 646–647, 662, 682–683
Particle model of light, 75–85
  and matter waves, 644
  and photons, 590–591
  difficulties with, 82–85
Pauli principle, 661–663
Pencils of light, 49
Penetration, 679–685
Period, 104, 242, 324
  of earth satellites, 246
  of simple pendulum, 249–250
  planetary, 265 (Table), 270
Periodic table of the elements, 660
Pendulum, 249–250
  damped, 403
  Foucault, 252–253
Periodic waves, 104
  See also Waves
Permutations, 390–391
Perpetual motion, 221
Perturbation theory, 272
Phase, 116, 121–123
  of light sources, 129–131, 569–570
Photocells, 46, 592–593
Photochemistry, 596
Photocurrent, 593–594
Photoelectric effect, 592–597, 632, 656
  Einstein-Planck relation, 596
Photographic film, 46
Photography, 155, 161, 592, 596
Photon, 591
  and atomic spectra, 628–634, 653
  diffraction of, 598
  energy of, 595–596, 653
  mass of, 609, 617–619
  momentum of, 609–612
  wave behavior of, 599–601
Photon emission, 557, 612, 625, 686–687
Photosensitive devices, 46
Pion, 616
Plane, inclined, 220
Planck, Max, 596
Planck's constant, 595–596, 644
Planetary atom, 467, 485–486
Planetary orbits, 265 (Table), 266, 488
  and gravitation, 269
  perturbations of, 272
Planetary system, 267–268
  and universal gravitation, 270–273
  dynamical approach to, 267–268
  early and Ptolemaic, 259–260
  of Copernicus, 260–262
  of Kepler, 263–267
  Of Tycho Brahe, 262
Planets, dimensions of, 265 (Table)
  discovery of, 273
  shape of, 38
  solar distance of, 267 (Table)
Plate, 435
Plato, 258
Position, 170
Position-time graph, 170, 176–185, 189
  Position vector, 196
Positron, 613
Potential, electric, 475–476
  measurement, of, 502
  unit of, 477–478
Potential barrier, 679
  and energy spread, 685–688
  penetration of, 680–685
Potential difference, 476, 498
  and simple circuits, 502–509
Potential energy, 316
  and internal energy, 376–377
  electric, 475–476
  gravitational, 328–332
  in a nuclear field, 486–487, 650
  in two-body interaction, 324–328
  molecular, 374
Potential hill, 475–476, 486–489
Potential well, 665, 679
  penetration of, 679–685
Potentiometer, 502
Pound, R. V., 618
Power, 473, 478–479
  in series-parallel circuits, 508–509
Power expansion, 565
Power generation, electric, 547
  nuclear, 615, 677
Power laws, 26–29, 34–35
Pressure, in a gas, 371–372
  of light, 81, 609–610
Principal rays, 154
Prisms, 66–70
  convergence of light by, 149–150
Probability, 388–394, 597–599
Probability wave, 643
  and lifetimes, 686
  leakage of, 681–686
Projectile motion, 206–207, 238–241, 330
Projectors, 155
Propagation of waves, 101–105, 134
  speed of, 103–104, 107–108
Proper time, 579
Proportionality constant, 25, 35

Proton, charge of, 456
  mass of, 465–466, 531
Proton absorption, 684–685
Ptolemy, 62, 260
Pulses, 88–96
  interference between, 114
  reflection and transmission of, 93–95, 102
  straight and circular, 100–102
  superposition of, 90–92, 115

Q

Quantum energy, 596
Quantum number, 647
Quantum theory, 644, 646, 651–668
  and energy spread, 686
  and wave model, 680–690
Quartz-crystal clock, 6
Quartz-fiber balance, 12

R

Radar, 436, 599
Radian, 351
Radiation, 47, 380
  and entropy, 417
  electromagnetic, 553–558
  See also Photon emission
Radio waves, 555–557, 599
Radioactive decay, age determination by, 4
  and measurement of mass, 12
  See also Nuclear disintegration
Radio-frequency spectra, 632
Rainbow, 68, 132
Ramsey, N. F., 611
Range finder, 8
Rays of light, 50
  principal, 154
Rebka, G. A., 618
Recoil, 282
Rectifier, 435
Red shift, 583
  gravitational, 618–619
Reflection, 45, 54–57
  and light pressure, 609–610
  and particle model, 76
  and thin films, 137–140
  angle of, 55
  internal, 66–68
  laws of, 55, 102
  of pulses, 93–95, 102, 134
Refraction, 46, 57–66
  and chromatic aberration, 161–162
  and particle model, 76–79
  and thin films, 139
  angle of, 58
  by prisms, 68–70, 149–150
  first law of, 59
  index of, 61–70, 63 (Table), 69 (Table)
  of waves, 105–107, 134
  Snell's law of, 62–63
Relativistic addition of velocities, 573, 577
Relativistic dynamics, 604–609
Relativistic kinetic energy, 606–608
Relativistic mass, 608
Relativistic momentum, 608–609
Relativity, 215, 275, 573–582
  special theory of, 578
Resistance, electrical, 424, 506
Resolution of images, 162–164
Rest energy, 607, 615
Rest mass, 608
Restoring forces, linear, 248–250, 323
Retrograde motion, 259
Reversible processes, 403, 407–409
Reversibility of light paths, 64–65, 151
Right-hand rule, 348, 516, 518, 521
Ripple tank, 100
Robinson, Sir Robert, 534
Rockets, 293–294, 332–334
Roemer, Olaf, 48
Rotation, and moment of inertia, 352–354
  of rigid bodies, 351
  uniform, 211, 241–244, 248
Rowland, Henry A., 516
Royds, J. T., 535
Rumford, Benjamin Thompson, Count, 376
Rutherford, Ernest, 456, 466, 485, 534, 625
Rydberg, J. R., 652
Rydberg constant, 652

S

Satellites, 220, 245–247
  and gravitational potential energy, 329, 331–334
  launching of, 293–294
  motion of, 345–346
Saturation current, 504
Savart, Felix, 524
Scalars, 199
Scaling, 23, 35–39
Scattering, alpha-alpha, 686
  alpha particle, 486–494, 677–679
  angle of, 488
  Compton, 611–612
  nuclear, 666–668
  of light, 76, 397
  proton-proton, 666–668
Screws, 8
Searchlights, 146
Semi-log plots, 32–34
Series circuits, 508–509
Shadows, 47, 108–110
Shells, 661
Shielding, electrical, 427–428
Similar figures, 27–28
Simple harmonic motion, 247–250, 323–324
Simultaneity, 576
Sky color, 397
Slope, 178–180
Smith, J. H., 580
Snell, W., 62
Snell's law, 62–63
  and particle model, 77
  for waves, 106–107
Soap films, 13, 137–140
Solar system, 265 (Table), 267 (Table)
  See also Planetary systems
Solenoid, 527–528
Space, measurement of, 6–11
Space charge, 505
Space fork, 2
Sparks, 504, 557
Spectra, atomic, 628–634, 652–660
  nuclear, 666–668
Spectral color, 70, 132 (Table), 13
Spectral elements, 148
Spectral lines, 625, 628–635
  formation of, 631
  of hydrogenlike atoms, 653–656, 655 (Table)
  widths of, 686–687
Spectrograph, mass, 532
Spectrometer, 654
Spectroscope, 133, 148
Spectrum, electromagnetic, 557
Specular reflection, 54–55, 76
Speed, 172, 201
  ultimate, 604–606
  See also Velocity
Speed of light, 48, 78, 83–84, 107, 215
  and electromagnetic radiation, 555–556
  and Newton's law, 229
  and relativistic dynamics, 604–609
  as ultimate speed, 606
  measurement of, 563–573
Spherical aberration, 160
Spin, 356–360
  and Pauli principle, 662
  in nuclear dissociation, 674–675
Springs, and potential energy, 320–324
  and simple-harmonic motion, 247
  interference on, 114–115
  waves on, 88–96
Square-well potential, 665, 679
  penetration of, 679–685
Standing waves, 141, 644–647
  and barrier penetration, 680–690
  and energy levels, 688–690
  in electron orbit, 652
Stars, fixed, 258, 261
  red shift of, 619
States and distributions, 389–394
Stellar distances, 30
Stirling's formula, 394
Stroboscope, 4–5
Sunlight, 600
Supernova, 292
Superposition principle, 91, 116
Sweep circuit, 436
Synchronization, 563–564, 575–577
Synchrotron, 556

T

Taylor, Sir Geoffrey I., 598
Telescopes, 147–149, 153–154
  diffraction in, 162–164
  refracting, 159–160
Television, 436, 592
Temperature, and electric resistance, 506
  and inelastic collisions, 315, 375, 397
  and internal energy, 375–378
  and molecular kinetic theory, 315, 373–374
  Kelvin, 368
Temperature rise, 315, 375, 378–379, 397–398
Thales of Miletus, 514
Thermionic emission, 434, 505–506
Thermodynamics, first law of, 379
  second law of, 379, 416
Thin films, 133, 137–140
Thomson, J. J., 471, 533
Threshold, photoelectric, 593, 632
Tides, 272
Time, dilation of, 578–582
  in moving frame of reference, 574–577
  measurement of, 3–6, 3 (Table), 563–564
  proper, 579
Tolerance, 14
Torque, 355–356
Trajectories, 344–346
Transparency, 44
Triangulation, 7–8, 13, 30
Tritium, 614–615
Troughs, 100–101
Turbine, 547
Twin paradox, 583–584
Tycho Brahe, 262–263

U

Ultraviolet radiation, 47, 557
Uncertainty, 600
Uranium fission, 615, 675–677, 683–684

V

Van de Graaff accelerator, 604–605
Vector acceleration, 204–212
Vectors, 196–204
Vector product, 348
Velocity, 171, 201
  and force applied, 222–228, 231
  angular, 351–352
  at constant acceleration, 187–189, 206–208
  average, 184
  constant, 176–180
  escape, 333
  in free fall, 206, 237–238, 328
  instantaneous, 180–185, 201, 206–207
  of projectiles, 208, 238–241
  relativistic addition of, 573–574
  terminal, 28
  variable, 173–176
  vector, 200–211
  See also Acceleration
Velocity-time graph, 171–175, 189
Visible light, 557, 596, 600
Vision, and distinct focusing, 155–157
  binocular, 51
Volt, 477
Voltmeter, 502–503

W

Wallis, John, 286
Watt, James, 331, 473
Watt (unit), 331, 473, 478
Waves, 87–97
  and photons, 599–601
  diffraction of, 108–110
  dispersion of, 107–108
  electromagnetic, 553–558
  interference between, 115–124
  in water, 100–110, 114–124
  matter, 639–647, 681–683
  periodic, 104
  refraction of, 105–107
  speed of propagation of, 103–104
  standing, 141, 644–647, 652, 680–684, 688–690
  superposition of, 90–93, 115, 646
  transmission and reflection of, 93–95
Wave mechanics, 643, 680–690
Wave model of light, 100–110, 127–141
  and ether, 562–571
  and matter waves, 644
  electromagnetic, 553–557
Wave-particle relationship, 644
  and alpha decay, 683–685
  in a "box", 646–647, 682–683
Wavelength, 104, 119–121
  and color, 132 (Table)
  and photoelectric effect, 593–594
  De Broglie, 638–644
Weight, 235
Well potential, 665, 679
  penetration of, 679–685
White dwarfs, red shift of, 619
White light, *See* Color
Work, 301, 303
  and kinetic energy, 304–305
  and potential energy, 322
  in a gravitational field, 332
  in isothermal gas expansion, 410–411
Wren, Sir Christopher, 286

X

X-ray absorption, 657–658, 661
X-rays, 430, 556, 558, 600

Y

Young, Thomas, 129
Young's experiment, 129, 137

Z

Zeeman, Pieter, 572

1 2 3 4 5 6 7 8 9 0